

CUBATURE METHOD TO SOLVE BSDEs: ERROR EXPANSION AND COMPLEXITY CONTROL

JEAN-FRANCOIS CHASSAGNEUX* AND CAMILO A. GARCIA TRILLOS†‡

ABSTRACT. We obtain an explicit error expansion for the solution of Backward Stochastic Differential Equations (BSDEs) using the cubature on Wiener spaces method. The result is proved under a mild strengthening of the assumptions needed for the application of the cubature method. The explicit expansion can then be used to construct implementable higher order approximations via Richardson-Romberg extrapolation. To allow for an effective efficiency improvement of the interpolated algorithm, we introduce an additional projection on sparse grids, and study the resulting complexity reduction. Numerical examples are provided to illustrate our results.

1. INTRODUCTION

Let $(\Omega, \mathbb{P}, \mathcal{F}, \mathbb{F})$ with $\mathbb{F} = (\mathcal{F}_t)_{t \in \mathbb{R}^+}$ be a filtered probability space satisfying the usual conditions. For some $T > 0$, we consider the solution of the Markovian Backward Stochastic Differential Equation

$$(1) \quad X_t = x_0 + \int_0^t b(s, X_s) ds + \int_0^t \sigma(s, X_s) dW_s,$$

$$(2) \quad Y_t = g(X_T) + \int_t^T f(s, X_s, Y_s, Z_s) ds + \int_t^T Z_s dW_s,$$

where W is an \mathbb{F} -Brownian Motion taking values in \mathbb{R}^r , X, Z are \mathbb{R}^d processes adapted to \mathbb{F} , Y is an \mathbb{F} -adapted process valued in \mathbb{R} , and $b : \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}^d$, $\sigma : \mathbb{R}_+ \times \mathbb{R}^d \rightarrow M(d, r)$, $f : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \rightarrow \mathbb{R}$ and $g : \mathbb{R}^d \rightarrow \mathbb{R}$ are Lipschitz function. In the sequel, we shall impose further regularity conditions for our theoretical analysis, see Assumptions 1.1 or 1.2 below.

An important property of the solution of a Markovian BSDE is that it can be represented as $Y_t = u(t, X_t)$ and $Z_t = v(t, X_t)$ for suitable functions u, v satisfying, at least in a viscosity sense, the PDE

$$(3) \quad \begin{cases} \partial_t u + \mathcal{L}u + f(\cdot, u, v) = 0 & \text{on } [0, T) \times \mathbb{R}^d, \\ u(T, \cdot) = g(\cdot) \end{cases},$$

where \mathcal{L} is the Dynkin operator associated to the diffusion X . Moreover, when sufficient regularity is available, we have that $v(t, X_t) := \partial_x u(t, X_t) \sigma(t, X_t)$.

Approximating (Y, Z) allows then to solve numerically and in a probabilistic way, the corresponding PDE for u . This has motivated in the past fifteen years an important literature on numerical methods for BSDEs. The main method to approximate (2) is a backward programming algorithm based on an Euler scheme, that has been introduced in [5] and [4, 35], see the references therein for early works. Since then, many extensions have been considered: high order schemes e.g. [9, 11], schemes for reflected BSDEs [2, 10], for fully coupled BSDEs [20, 3], for quadratic BSDEs [8] or McKean-Vlasov BSDEs

*Laboratoire de Probabilités et Modèles Aléatoires, Université Paris Diderot. chassagneux@math.univ-paris-diderot.fr

†Department of Mathematics, University College London. camilo.garcia@ucl.ac.uk.

‡The author's research was partially supported by OpenGamma Ltd.

[13, 12]. It is also important to mention that, quite recently, very promising probabilistic forward methods have been introduced to approximate (2) [6] or directly the non-linear parabolic PDE (3) [26]. The backward algorithm approximating (2) requires, to be fully implementable, a good approximation of (1) and its associated conditional expectation operator. Various methods have been developed, see e.g. [19, 31, 23] and we will focus here on the cubature on Wiener spaces introduced in [29]. Broadly speaking, the method considers the space of continuous paths in \mathbb{R}^d on the interval $[0, T]$ ($C([0, T], \mathbb{R}^d)$) to define a finite probability \mathbb{Q} that approximates the Wiener law \mathbb{P} . As we explain briefly in Section 2.2, see a full account in [29], this approximation is chosen to match the expectation of iterated integrals.

The paper [16] pioneered the use of the cubature method to solve BSDEs. Essentially, the algorithm estimates the value of the field u on the points on the support of the cubature approximating law, thus giving an approximation scheme for the solution of the BSDE (2). By its nature, this algorithm can be easily used to implement second order discretization schemes as in [9, 17], and applied in the context of McKean-Vlasov BSDEs as in [13]. The cubature algorithm has been studied under a set of assumptions that guarantee sufficient regularity for the field (for example, smooth coefficients for the forward equation and the generator of the backward equation, Lipschitz regularity on the boundary condition plus a structural condition of the type UFG, see Section 1.1 below).

In this work, we want to study acceleration methods for the Euler approximation of BSDEs, of the same kind as those proved in the linear case ($f = 0$) [33] (see also [22] for the study of the discrete-time error only in the non-linear case). We show that under a very mild strengthening of the assumptions, there exists an explicit error expansion for the weak approximation of the BSDE system given by equations (1) and (2) using the cubature on Wiener spaces method. The explicit expansion exposes the dependence of the error approximation with respect to the general features of the coefficients of the system and the test function. Moreover, it opens the possibility to increase the rate of convergence by using Richardson-Romberg extrapolation techniques. However, to effectively improve the efficiency of the algorithm, we need to analyse and improve the complexity growth of the approximation technique. In this work, we consider a technique based on sparse grids.

Finally, note that although in the paper we limit our study to a first order scheme, a similar study can be carried out to analyze schemes with higher convergence rate as the ones proposed in [11].

We now present the setup of our work and the numerical schemes we study in the sequel, and give an overview of the main results of the paper.

1.1. Main Assumptions. Let us rewrite the process X in (1) in its Stratonovich form, that is, let

$$X_t = x_0 + \int_0^t \bar{b}(s, X_s) ds + \int_0^t \sigma(s, X_s) \circ dW_s$$

where $\bar{b} : \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ has i -th component defined by

$$(4) \quad \bar{b}_i(t, x) = b_i(t, x) - \frac{1}{2} \sum_{j=1}^r \sum_{k=1}^d \sigma_{k,j}(t, x) \partial_{x_k} \sigma_{i,j}(t, x).$$

We work under two sufficient assumptions to guarantee the regularity needed for the cubature method to be effective.

Assumption 1.1. *Let $M > 0$.*

- *On the forward coefficients*
 - i. $\bar{b}, \sigma_{\cdot, j} \in C_b^{M+1}(\mathbb{R}^+ \times \mathbb{R}^d, \mathbb{R}^d)$ for all $j = 1, \dots, d$;
- *On the backward coefficients*
 - i. $g \in C_b^{M+1}(\mathbb{R}^d, \mathbb{R})$

$$ii. f \in C_b^{M+1}(\mathbb{R}^+ \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d, \mathbb{R}).$$

Assumption 1.2. Let $M > 0$. There exist $\alpha > 0$, such that

- On the forward coefficients:
 - i. $\bar{b} \in C_b^{M+\alpha+1}(\mathbb{R}^+ \times \mathbb{R}^d, \mathbb{R}^d)$ and $\sigma_{.,j} \in C_b^{M+\alpha+2}(\mathbb{R}^+ \times \mathbb{R}^d, \mathbb{R}^d)$ for all $j = 1, \dots, d$;
 - ii. UFG condition of order α (see Definition 1.1. in [14])
- On the backward coefficients
 - i. g is Lipschitz continuous.
 - ii. $f \in C_b^M(\mathbb{R}^+ \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d, \mathbb{R})$

The existence of a classical solution to the PDE (3) is assured under either of Assumptions 1.1 or 1.2.

1.2. Forward scheme. We define a stochastic process on $C_{bv}^0([0, T], \mathbb{R}^d)$ – the space of continuous functions with bounded variation – as follows. Let $\hat{X} : \hat{\Omega} \rightarrow C_{bv}^0([0, T], \mathbb{R}^d)$ be given as the solution of

$$(5) \quad \hat{X}_t(\hat{\omega}) := x_0 + \int_0^t \bar{b}(s, \hat{X}_s(\hat{\omega})) ds + \int_0^t \sigma(s, \hat{X}_s(\hat{\omega})) d\hat{\omega}_s.$$

The integrals in the previous definition are taken in the Riemann-Stieltjes sense, which is possible since we have assumed that the paths $\hat{\omega}$ are of bounded variation. Let us also define a conditioned form of this stochastic process, given by

$$(6) \quad \hat{X}_t^{s,x}(\hat{\omega}) := x + \int_s^t \bar{b}(r, \hat{X}_r^{s,x}(\hat{\omega})) dr + \int_s^t \sigma(r, \hat{X}_r^{s,x}(\hat{\omega})) d\hat{\omega}_r.$$

We take as weak approximation of the process X the process \hat{X} under a finite cubature measure $\hat{\mathbb{Q}}$ (see the precise definition in Section 2.2 below). In other words, we consider a random process in a finite space obtained by solving a finite number of ODEs. For practical implementation, the resulting discrete measure is built as a tree. This allows an easy computation of conditional expectations, a property that is of paramount importance to solve the Backward component that we introduce below.

The precision of approximation provided by the cubature method is given in terms of its order: it quantifies the degree of iterated integrals that can be perfectly computed in expectation under the cubature measure. Roughly speaking, this is analogous to the maximal degree of polynomials perfectly approximated by quadrature rules in finite dimensional spaces.

1.3. Backward scheme. As mentioned before, in the Markovian setting it is possible to express the solution to the BSDE equation (2) in terms of the so-called *decoupling field*, that is, applications $u : \mathbb{R}_+, \mathbb{R}^d \rightarrow \mathbb{R}$ and $v : \mathbb{R}_+, \mathbb{R}^d \rightarrow \mathbb{R}^d$ defined by

$$u(t, x) = Y_t^{t,x}; \quad v(t, x) = Z_t^{t,x}.$$

For $\gamma \geq 1$, we consider a time grid of the form

$$(7) \quad t_i = T \left[1 - \left(1 - \frac{i}{n} \right)^\gamma \right]$$

and we set $h_i := t_{i+1} - t_i$ for $i = 0, \dots, n$. We study a cubature based BTZ scheme defined by

- (i) Terminal condition is $(\hat{u}_n, \hat{v}_n) = (g, 0)$

(ii) Transition from step $i + 1$ to step i given by

$$\begin{aligned}\hat{u}_i(x) &= \mathbb{E}^{\hat{\mathbb{Q}}} \left[\hat{u}_{i+1}(\hat{X}_{t_{i+1}}^{t_i, x}) + h_i f(x, \hat{u}_i(x), \hat{v}_i(x)) \right], \\ \hat{v}_i(x) &= \mathbb{E}^{\hat{\mathbb{Q}}} \left[\hat{u}_{i+1}(\hat{X}_{t_{i+1}}^{t_i, x}) \frac{\Delta \hat{\omega}_i}{h_i} \right],\end{aligned}$$

where $\Delta \hat{\omega}_i = \hat{\omega}_{t_{i+1}} - \hat{\omega}_{t_i}$, and $\mathbb{E}^{\hat{\mathbb{Q}}}$ is the expectation with respect to the cubature measure. We then define $\hat{Y}_i = \hat{u}_i(\hat{X}_{t_i})$ and $\hat{Z}_i = \hat{v}_i(\hat{X}_{t_i})$.

1.4. Main results.

1.4.1. *Error expansion.* Our first result, Theorem 1.3 extends the results of [33] on the Euler scheme for weak approximations of SDEs to the case where the underlying numerical method is not the Monte Carlo method but the cubature on Wiener spaces.

Theorem 1.3 (Forward error expansion). *Set $m \geq 3$. Suppose that Assumption 1.1 (resp. 1.2) holds with $M \geq m + 2$, and take $\hat{\mathbb{Q}}$ to be a cubature measure from a cubature formula of order m on a uniform (resp. decreasing) step grid with $\gamma = 1$ (resp. $\gamma > m - 1$).*

Then, there is a constant K such that, for all $i < n$,

$$(8) \quad \left| \mathbb{E}^{\hat{\mathbb{Q}}} \left[g(X_T^{t_i, x}) \right] - \mathbb{E} \left[g(X_T^{t_i, x}) \right] \right| \leq K n^{-\frac{m-1}{2}}.$$

Moreover, if $M \geq m + 3$ in Assumption 1.1 (resp. 1.2), then

$$\left| \mathbb{E}^{\hat{\mathbb{Q}}} \left[g(X_T^{t_i, x}) \right] - \mathbb{E} \left[g(X_T^{t_i, x}) \right] - n^{-\frac{m-1}{2}} \Psi_T^{lin}(t_i, x) \right| \leq K' n^{-\frac{(m+1)\wedge \gamma}{2}},$$

where the coefficient Ψ_T^{lin} is given in Definition 2.9 below.

Our main result, see Theorem 1.4 below, carries out a similar analysis taking into account the non-linearity associated to the generator function f in the formulation of BSDEs. Let us stress that this analogous result is obtained on a *completely implementable* scheme to solve BSDEs, that is, we include the analysis on the conditional expectation approximation.

Theorem 1.4 (Backward error expansion). *Suppose that Assumption 1.1 (resp. 1.2) holds with $M \geq 9$, and take $\hat{\mathbb{Q}}$ to be a cubature measure from a cubature formula of order $m \geq 3$ on a uniform (resp. decreasing) step grid with $\gamma = 1$ (resp. $\gamma > m - 1$). Then, for all $i < n$,*

$$(9) \quad \left| \hat{u}_{t_i}(x) - u(t_i, x) + n^{-1} \Psi_T^{nl}(t_i, x) \right| \leq K n^{-2},$$

where the coefficient Ψ_T^{nl} is given in Definition 3.9 below.

1.4.2. *Complexity reduction.* With the above error expansion at hand, one can implement a Romberg-Richardson method to increase the precision of the approximation profiting from the regularity of the value function u . This is done by using an order 1 method, here a modified Euler scheme, which is found often to be easier to implement than a high order scheme and may exhibit also better numerical stability property (see [8] for a study of numerical stability in the case of BSDEs approximation). With an increased precision, one can hope to lower the numerical complexity of the method. However, this is not the case here if one simply uses the cubature tree to compute the conditional expectation.

Indeed, the main drawback of the cubature method is its complexity growth with respect to the number of time discretization steps on the approximation. In the case of approximation of expectations, complexity can be controlled using reduction techniques as high order recombination [28] and TBBA (Tree Based Branching Algorithm) [15].

To make the extrapolation method worth implementing in practice, we introduce an extra projection on a finite grid. In order to reduce the complexity further, we use a *sparse*

grid approach. Note that, the use of *sparse grid* has already been suggested in the context of BSDEs approximation in [34], but the forward approximation method used in this paper is different. Moreover, the theoretical analysis is done under restrictive condition on the generator f , the forward process being restricted to be the Brownian motion itself.

To simplify the discussion, we only examine the case where the diffusion propagates in all \mathbb{R}^d . We then strengthen the regularity conditions on the data of the problem considering instead the uniform Hörmander condition, that is

Assumption 1.5 (Uniform Hörmander condition). *Let \mathcal{V}_k be defined by*

$$\mathcal{V}_0 = \{V^{(i)}; i > 0\}; \quad \mathcal{V}_k = \mathcal{V}_k \cup \{[V, V^{(i)}]; V \in \mathcal{V}_k, i \geq 0\}$$

where $[\cdot, \cdot]$ denotes the Lie product. The uniform Hörmander condition of order M is satisfied if

$$\inf_{x, \xi \in \mathbb{R}^d, |\xi|=1, v \in \mathcal{V}_M} \{(v(x) \cdot \xi)^2\} > 0.$$

This condition is stronger than the UFG condition.

As mentioned previously, our implemented algorithm uses the concept of *sparse grid* that we introduce briefly now, see [7] for a detailed account on this topic.

Let $A = \prod_{i=1}^d [a_i, b_i]$ be a hypercube in \mathbb{R}^d . Given a multi-index $\mathbf{l} \in \mathbb{N}^d$ we define a vector of grid sizes

$$\delta_{\mathbf{l}}(A) := (\delta_{l_1}, \dots, \delta_{l_d}); \quad \text{where } \delta_{l_i} = (b_i - a_i) \cdot 2^{-l_i} \text{ for } i = 1, \dots, d.$$

This set of distances defines a grid with nodes denoted by

$$\tilde{\mathbf{x}}_{\mathbf{l}, \mathbf{j}}(A) := (a_1 + j_1(b_1 - a_1)\delta_{l_1}, \dots, a_d + j_d(b_d - a_d)\delta_{l_d}), \quad \text{for } \mathbf{0} \leq \mathbf{j} \leq \mathbf{2}^{\mathbf{l}}.$$

We denote $\tilde{x}_{l_i, j_i}^i := a_i + j_i(b_i - a_i)\delta_{l_i}$, $1 \leq i \leq d$, and observe that $\tilde{\mathbf{x}}_{\mathbf{l}, \mathbf{j}}(A) = (\tilde{x}_{l_i, j_i}^i)_{1 \leq i \leq d}$. By setting

$$\phi(x) := \begin{cases} 1 - |x| & \text{if } x \in [-1, 1] \\ 0 & \text{otherwise} \end{cases},$$

we can define an associated set of nodal basis functions given by

$$(10) \quad \phi_{\mathbf{l}, \mathbf{j}}(x; A) := \prod_{i=1}^d \phi\left(\frac{x_i - \tilde{x}_{l_i, j_i}^i}{\delta_{l_i}}\right).$$

We consider the sparse grid nodal space of order p defined by

$$\mathcal{V}_p(A) := \text{span} \{\phi_{\mathbf{l}, \mathbf{j}}; (\mathbf{l}, \mathbf{j}) \in \mathcal{I}_p(A)\},$$

where

$$(11) \quad \mathcal{I}_p(A) := \left\{(\mathbf{l}, \mathbf{j}) : 0 \leq \sum_{i=1}^d l_i \leq p; \quad \mathbf{0} \leq \mathbf{j} \leq \mathbf{2}^{\mathbf{l}}; \right. \\ \left. (l_i > 0 \text{ and } j_i \text{ is odd}) \text{ or } (l_i = 0), \text{ for } i = 1, \dots, d\right\}.$$

For a function $\psi : A \rightarrow \mathbb{R}$ with support in A , we define its associated \mathcal{V}_p -interpolator by

$$(12) \quad \pi_{\mathcal{V}_p}^A[\psi](x) := \sum_{(\mathbf{l}, \mathbf{j}) \in \mathcal{I}_p(A)} \theta_{\mathbf{l}, \mathbf{j}}(\psi; A) \phi_{\mathbf{l}, \mathbf{j}}(x; A)$$

where the operator $\theta_{\mathbf{l}, \mathbf{j}}$ can be defined recursively in terms of r , the dimension of \mathbf{l} , by:

$$(13) \quad \theta_{\mathbf{l}, \mathbf{j}}(\psi; A) = \begin{cases} \psi(\tilde{x}_{\mathbf{l}, \mathbf{j}}); & r = 0 \\ \theta_{\mathbf{l}-, \mathbf{j}-}(\psi(\cdot, \tilde{x}_{l_r, j_r}^r); A-); & l_r = 0 \\ \theta_{\mathbf{l}-, \mathbf{j}-}(\psi(\cdot, \tilde{x}_{l_r, j_r}^r); A-) - \frac{1}{2}\theta_{\mathbf{l}-, \mathbf{j}-}(\psi(\cdot, \tilde{x}_{l_r, j_r-1}^r); A-) \\ \quad - \frac{1}{2}\theta_{\mathbf{l}-, \mathbf{j}-}(\psi(\cdot, \tilde{x}_{l_r, j_r+1}^r); A-); & l_r > 0 \end{cases}$$

where, for a hypercube $A = \prod_{i=1}^d [a_i, b_i]$, $A- := \prod_{i=1}^{d-1} [a_i, b_i]$ and for a multi-index \mathbf{k} with dimension $r \geq 1$, $\mathbf{k}- = (k_1, \dots, k_{r-1})$.

Finally, let us introduce a useful notation. For a bounded set H , we denote by

$$(H)_p^\vee = \{\tilde{\mathbf{x}}_{\mathbf{l}, \mathbf{j}}((H)^\blacksquare); (\mathbf{l}, \mathbf{j}) \in \mathcal{I}_p((H)^\blacksquare)\}$$

where $(H)^\blacksquare$ is the minimal hypercube that contains H , i.e.

$$(H)^\blacksquare := \bigcap \left\{ A : H \subset A, A = \prod_{i=1}^d [a_i, b_i] \text{ for } \mathbf{a}, \mathbf{b} \in \mathbb{R}^d \right\}.$$

With the sparse representation at hand, we modify the backward scheme and introduce first the set of points where the function needs to be approximated. Let

$$(14) \quad D_i = \begin{cases} \{x_0\} & \text{if } i=0 \\ \left(\bigcup_{x \in D_{i-1}} \text{supp} \left[\hat{X}_{t_i}^{t_{i-1}, x} \right] \right)_{p_i}^\vee & \text{if } i=1, \dots, n-1 \end{cases},$$

so that D_i is a set of points in a sparse grid of order p_i . This sparse grid is constructed on the minimal hypercube that contains the diffusion started at D_{i-1} at time t_{i-1} . We denote by \mathcal{D} the union of all the points in the grids D_i , namely

$$(15) \quad \mathcal{D} := \bigcup_{i=0}^{n-1} \{t_i\} \times D_i$$

which forms a finite grid of $[0, T) \times \mathbb{R}^d$.

Having this set, for a sequence of values (p_1, \dots, p_{n-1}) with $p_i > d-1$, we define the sparse-cubature backward approximation by

$$(16) \quad \check{u}_i(x) = \pi_{\mathcal{V}_{p_i-d+1}}^{D_i} \left[\left(\mathbb{E}^{\hat{\mathbb{Q}}} \left[\check{u}_{i+1} \left(\hat{X}_{t_{i+1}}^{t_i, \cdot} \right) \right] \right) \right](x) + h_i f(x, \check{u}_i(x), \check{v}_i(x))$$

$$(17) \quad \check{v}_i(x) = \pi_{\mathcal{V}_{p_i-d+1}}^{D_i} \left[\left(\mathbb{E}^{\hat{\mathbb{Q}}} \left[\check{u}_{i+1} \left(\hat{X}_{t_{i+1}}^{t_i, \cdot} \right) \frac{\Delta \hat{w}_i}{h_i} \right] \right) \right](x)$$

for $x \in D_i^\blacksquare$. The terminal condition is set to $(\check{u}_n, \check{v}_n) = (g, 0)$.

In practice, the computational effort to obtain $(\check{u}_i, \check{v}_i)_i$ is proportional to the number of points in the grid \mathcal{D} . But, let us insist on the fact that, our approximation is then defined and available, without loss of precision, on the bigger space $\bigcup_{i=0}^{n-1} \{t_i\} \times D_i^\blacksquare$.

Our main result for this part concerns the complexity of the method and the fact that it can be improved by using an extrapolation.

We are interested here on the case when $m = 3$ and we fix a symmetric cubature formula \mathbb{Q} with the given order size κ . For convenience, set

$$\chi_n^\gamma := (n, (t_0, \dots, t_n)),$$

where (t_i) are given by (7) with parameter γ , and denote by $\check{u}^{\chi_n^\gamma}$, the solution of the cubature and sparse grid scheme, constructed with parameters χ_n^γ and (p_1, \dots, p_{n-1}) is defined specifically, see Proposition 4.5 for details.

Theorem 1.6. *Suppose that Assumption 1.5 and either Assumption 1.1 or 1.2 hold with $M > (2d+1) \vee 9$. Let $\Delta > 0$ be fixed. Assume that $m = 3$ and let $\gamma = 1$ if Assumption 1.1 holds and $\gamma = 3$ otherwise.*

i. *Complexity: Let*

$$N = \inf\{z \in \mathbb{N} : |\tilde{u}_0^{\chi_z}(x) - u(0, x)| < \Delta\},$$

that is, the minimal number of steps needed to solve the problem with accuracy Δ .

Then, for every $\epsilon > 0$, there exists $K_\epsilon > 0$ (possibly depending on d) such that the complexity (measured in number of nodes) associated to \tilde{u}^{χ_N} is bounded by

$$\#\mathcal{D} \leq \begin{cases} K_\epsilon \Delta^{-(3+\epsilon)} & \text{under Assumption 1.1} \\ K_\epsilon \Delta^{-[(3\vee(d+7/6))+\epsilon]} & \text{under Assumption 1.2} \end{cases}.$$

ii. *Complexity of Extrapolation: Let*

$$(18) \quad \hat{t}_i = \begin{cases} t_j & \text{if } i = 2j \\ \frac{t_j + t_{j+1}}{2} & \text{if } i = 2j+1 \end{cases}$$

and

$$\hat{\chi}_{2n}^\gamma := (2n, (\hat{t}_0, \dots, \hat{t}_{2n})).$$

Let

$$\hat{N} = \inf\{z \in \mathbb{N} : |2\tilde{u}_0^{\hat{\chi}_{2z}^\gamma}(x) - \tilde{u}_0^{\chi_z^\gamma}(x) - u(0, x)| < \Delta\},$$

that is, the minimal number of steps needed to solve the problem with accuracy Δ , when using extrapolation with the refined grid presented before.

Then, for every $\epsilon > 0$, there exists $K_\epsilon > 0$ (possibly depending on d) such that the complexity (measured in number of nodes) associated to $2\tilde{u}_0^{\hat{\chi}_{2z}^\gamma}(x) - \tilde{u}_0^{\chi_z^\gamma}(x)$ is bounded by

$$\#\mathcal{D} \leq \begin{cases} K_\epsilon \Delta^{-(2+\epsilon)} & \text{under Assumption 1.1} \\ K_\epsilon \Delta^{-[(2\vee\frac{1}{2}(d+7/6))+\epsilon]} & \text{under Assumption 1.2} \end{cases}.$$

It is worth comparing this result to the complexity exposed in [24]: We see that our method in the smooth case has a better complexity (the order is not impacted by d). This comes directly from the use of *sparse grids* approximation.

Finally, let us mention, that if one is only interested in the approximation of the initial value $u(0, x)$, the complexity can be further reduced by excluding the zero-level points in the sparse grid approximation. The impact is on the constant K above, not the order, as precisely explained in Section 4.2.1.

1.4.3. *Numerical results.* We demonstrate the efficiency of our approximation scheme and, in particular, the gain in complexity coming from the Romberg-Richardson method. The system, we solve numerically, satisfies all of the underlying assumptions, and reads as follows:

- Forward equation: $X = W$, a d -dimensional Brownian Motion.
- Backward equation:

$$Y_t = g(W_1) + \int_t^1 f(Y_s, Z_s) ds - \int_t^1 Z_s dW_s$$

with $f : \mathbb{R} \times \mathbb{R}^d \rightarrow \mathbb{R}$ given by

$$f(y, z) = \left(y - \frac{2+d}{2d}\right) \sum_{\ell=1}^d z_\ell$$

and

$$g(x) = \frac{k(1, x)}{1 + k(1, x)}, \quad \text{with } k(1, x) = \exp\left(1 + \sum_{\ell=1}^d x_\ell\right).$$

We use a cubature formula in dimension d of order $m = 3$. It is defined, for $j = 1, \dots, 2d$ by $p_j = (2d)^{-1}$ and $\omega_j = (-1)^j e_{[j/2]}$, where e_i is the i th canonical basis element (see [25]).

The previous system can be solved analytically. In particular, we get

$$Y_0 = \frac{k(0, \mathbf{0})}{1 + k(0, \mathbf{0})} = \frac{1}{2}.$$

We look at the results of the algorithm and its extrapolated version, when we use the sparse grid implementation. The rate of convergence is shown in Figure 1 (Left). The original algorithm shows the expected rate of convergence. The extrapolated one converges even faster, showing there is possibly an extra cancellation for the next order term.

In Figure 1 (Right), an illustration of the rate of convergence and complexity of the scheme in terms of the time complexity of the algorithm is shown. We can see that there is an effective reduction on the overall time to solve the problem with a given error.

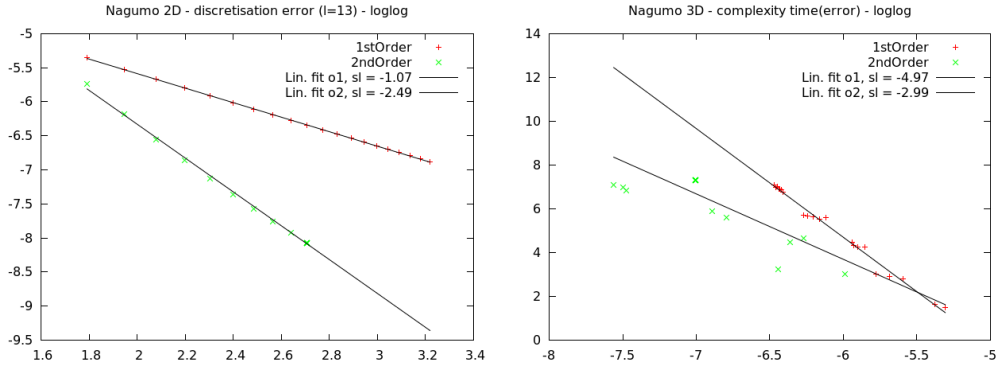


FIGURE 1. Left: Log of the error of the scheme in terms of the log of the number of steps (n). The expected rate of convergence is observed for the original algorithm and a better than expected for the extrapolated one. (Example in dimension 2). Right: Log-complexity time (in seconds) as a function of log error (Example in dimension 3). Data as in the text.

The rest of the paper is organised as follows. In Section 2, we study the approximation of the BSDE in the linear case, associated to the cubature method. In Section 3, we obtain an error expansion for the general case $f \neq 0$. In Section 4, we focus on the complexity reduction via the use of sparse grids and extrapolation methods. Finally, The Appendix collects some useful results on sparse grids and on integral approximation by Riemann sums.

2. CONVERGENCE ANALYSIS FOR THE FORWARD PROCESS

We first provide an error extrapolation for the BSDEs when $f = 0$. This result is new in the context of cubature method and terminal condition with Lipschitz only regularity.

2.1. Notation. In this section, we highlight the notation we use in all the document and that might not be considered completely standard.

2.1.1. Multi-indices. Multi-indices allow to easily manage differentiation and integration in several dimensions. Let us consider the set of multi-indices

$$(19) \quad \mathcal{M} = \{\emptyset\} \cup \bigcup_{l \in \mathbb{N}^*} \{0, 1, \dots, d\}^l,$$

where \emptyset refers to the zero-length multi-index. We define “ $*$ ” to be the natural concatenation operator, and we consider some norms in \mathcal{M} . For $\beta = (\beta_1, \dots, \beta_l)$:

$$|\beta|_p = \sum_{i=1}^l |\beta_i|^p, \text{ for } p \in \mathbb{N}; \quad |\beta|_\infty = \max_i |\beta_i|,$$

and

$$\|\beta\| := (\#\{i : \beta_i = 0\}) + |\beta|_0.$$

Naturally $|\emptyset|_p = \|\emptyset\| = 0$. For every $\beta \neq \emptyset$, we set $-\beta := (\beta_2, \dots, \beta_l)$; $\beta- := (\beta_1, \dots, \beta_{l-1})$ and for $1 \leq i < j \leq |\beta|_0$ $\beta_{i:j} := (\beta_i, \dots, \beta_j)$. We set $\beta^{>0}$ the multi-index obtained by deleting the zero components of β .

We refer to the set of multi-indices of degree at most l denoted by $\mathcal{A}_l := \{\beta \in \mathcal{M} : \|\beta\| \leq l\}$, and to its *frontier set* $\partial\mathcal{A} := \{\beta \in \mathcal{M} \setminus \mathcal{A} : -\beta \in \mathcal{A}\}$. It is readily seen that $\partial\mathcal{A}_l \subset \mathcal{A}_{l+2} \setminus \mathcal{A}_l$.

2.1.2. Stratonovich differential operators. As we show in section 2.2 (see also [29]), the cubature on Wiener spaces method uses the algebraic structure of the differential operators associated to the Stratonovich integral. We introduce then a special notation for the Stratonovich operators associated to the SDE (1) and their iterated actions.

Let $\psi : \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}$, then we define

$$\begin{aligned} V^\emptyset \psi(t, x) &= \psi(t, x) \\ V^{(0)} \psi(t, x) &= \sum_{i=1}^d \bar{b}_i \partial_{x_i} \psi(t, x) + \partial_t \psi(t, x) \\ V^{(j)} \psi(t, x) &= \sum_{i=1}^d \sigma_{i,j} \partial_{x_j} \psi(t, x); j = 1, \dots, d \end{aligned}$$

and for any multi-index $\beta \neq \emptyset$ with $|\beta| > 1$,

$$V^\beta(t, x) = V^{(\beta_1)}[V^{-\beta} \psi](t, x)$$

2.1.3. Space of differentiable functions. We denote by \bar{C}_b^m the class of differentiable functions $\psi : \mathbb{R}^d \rightarrow \mathbb{R}$ with bounded derivatives $V^\beta \psi$ for every $\beta \in \mathcal{A}_m$, i.e. functions whose semi-norm defined by

$$(20) \quad \|\psi\|_{m,\infty} := \sup_{\beta \in \mathcal{A}_m} |V^\beta \psi|_\infty.$$

is finite. In the definition, $|\cdot|_\infty$ stands for the usual maximum norm.

For convenience, for a fixed discretization grid, and $\psi : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$, we write

$$(21) \quad \|\psi\|_{i;m,\infty} := \sup_{t \in [t_i, t_{i+1}]} \|\psi(t, \cdot)\|_{m,\infty}.$$

2.1.4. Ito differential operators. Let $\psi : \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}$, then we define

$$L^{(0)} \psi(t, x) = \partial_t \psi(t, x) + \sum_{i=1}^d b_i \partial_{x_i} \psi(t, x) + \sum_{i=1}^d \sum_{j=1}^d \sum_{k=1}^d \sigma_{ik} \sigma_{kj} \partial_{x_i x_j} \psi(t, x).$$

Note that this operator can be written in terms of the Stratonovich differential operators we introduced, as follows

$$L^{(0)} = V^{(0)} + \sum_{j=1}^d V^{(j,j)}.$$

2.1.5. Iterated integrals. Let ω and ζ be two functions in $C_{\text{bv}}([0, T], \mathbb{R}^d)$, the space of continuous functions with bounded variation defined from $[0, T]$ to \mathbb{R}^d . Let us define the iterated integral of ζ with respect to ω by

$$I_{s,t}^\beta(\zeta, \omega) := \int_{s < t_1 < \dots < t_{|\beta|} \leq t} \zeta_{t_1} d\omega^{\beta_1}(t_1) \cdots d\omega^{\beta_{|\beta|}}(t_{|\beta|})$$

where ω^i is the i -th component of ω , and we fix by convention $\omega^0(t) := t$. In the following, we write $I_{s,t}^\beta(\omega) = I_{s,t}^\beta(1, \omega)$.

We introduce a similar notation to represent iterated integrals with respect to the Brownian motion. Indeed, let X be an adapted process. We set

$$J_{s,t}^\beta(X) := \int_{s < t_1 < \dots < t_{|\beta|} < t} X_{t_1} \circ dW_{t_1}^{\beta_1} \cdots \circ dW_{t_{|\beta|}}^{\beta_{|\beta|}}$$

where the notation \circ denotes the Stratonovich integral, and we keep the convention $W_t^0 = t$. Moreover, set $J_{s,t}^\beta := J_{s,t}^\beta(1)$.

2.1.6. Conditional process. In what follows, for any $0 \leq s \leq t < T$, we denote by $X_t^{s,x}$ the process X_t conditioned to be equal to x at time $s < t$, i.e.

$$X_t^{s,x} = x + \int_s^t b(r, X_r^{s,x}) dr + \int_s^t \sigma(r, X_r^{s,x}) dW_r$$

2.1.7. Operators. We consider now a family of operators associated with the SDE (1) and the process (5).

Definition 2.1 (Operators). *We denote by P the operator over measurable functions associated to the diffusion X defined, for any measurable function $\psi : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$, by*

$$P_{s,t}\psi(s, x) := \mathbb{E}[\psi(t, X_t^{s,x})].$$

Similarly, we define by Q the analogous cubature operator given by

$$Q_{s,t}\psi(s, x) := \mathbb{E}^{\hat{\mathbb{Q}}}[\psi(t, \hat{X}_t^{s,x})].$$

By a slight abuse of notation, we use this notation also for one-parameter functions, that is, if $\psi : \mathbb{R}^d \rightarrow \mathbb{R}$, $P_{s,t}\psi$ (respectively $Q_{s,t}\psi$) denotes the operator applied to the function $\bar{\psi} : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ such that $(t, x) \rightarrow \bar{\psi}(t, x) := \psi(x)$.

Note that, by Definition 2.1 and (6), we have

$$Q_{s,t}Q_{t,u}\psi(s, x) = \mathbb{E}^{\hat{\mathbb{Q}}}[Q_{t,u}\psi(t, \hat{X}_t^{s,x})] = \mathbb{E}^{\hat{\mathbb{Q}}}[\psi(u, \hat{X}_u^{t, \hat{X}_t^{s,x}})] = \mathbb{E}^{\hat{\mathbb{Q}}}[\psi(u, \hat{X}_u^{s,x})] = Q_{s,u}\psi(s, x).$$

2.2. Cubature on Wiener spaces. The cubature measure on Wiener spaces was introduced in [29], as a tool to construct weak approximations of functionals of the Brownian motion. It generalises the quadrature method to an infinite dimensional space: It aims at approximating the Wiener measure restricted to a time interval $[0, T]$ with a finite probability defined on $C([0, T], \mathbb{R}^d)$. As in the quadrature method, this approximation consists in preserving the exact value of the expectation of some basic functionals that will play a similar role as the one played by polynomials in finite dimensions.

Definition 2.2 (Cubature formula [29]). *Let m be a natural number. An m -cubature formula on the Wiener space $C^0([0, 1], \mathbb{R}^d)$ is a probability measure \mathbb{Q} with finite support on $C_{\text{bv}}^0([0, 1], \mathbb{R}^d)$ (continuous functions and bounded variation starting in 0) such that the expectation of the iterated Stratonovich integrals of degree m under the Wiener measure and under the cubature measure \mathbb{Q} are the same, i.e., for all multi-index $\beta \in \mathcal{A}_m$*

$$\mathbb{E} [J_{0,1}^\beta] = \mathbb{E}^\mathbb{Q} [I_{0,1}^\beta(\omega)] := \sum_{j=1}^{\kappa} \theta_j [I_{0,1}^\beta(\omega_j)]$$

where $\{\omega^1, \dots, \omega^\kappa\}$ and $\theta_1, \dots, \theta_\kappa$ are respectively the support and weights of the finite measure \mathbb{Q} .

Examples of cubature formulas of order 3 and 5 are known for arbitrary dimensions. Higher order cubature methods (up to order 11) are also given for small dimensions, see e.g. [29, 25].

Definition 2.2 may be extended to an m -cubature formula on the Wiener space $C^0([0, t], \mathbb{R}^d)$, for an arbitrary $t > 0$. Indeed, the rescaling properties of the Brownian motion imply that $\omega^1, \dots, \omega^\kappa$ and $\theta_1, \dots, \theta_\kappa$ form an m -cubature formula on $C_{\text{bv}}^0([0, 1], \mathbb{R}^d)$ if and only if $t^{1/2}\omega^1, \dots, t^{1/2}\omega^\kappa$ and $\theta_1, \dots, \theta_\kappa$ form an m -cubature formula on $C_{\text{bv}}^0([0, t], \mathbb{R}^d)$. This justifies the choice of giving the definition on the interval $[0, 1]$.

Definition 2.3 (Symmetric cubature formula). *We say that a cubature formula is symmetric if for any path $\omega^* \in \text{supp}(\mathbb{Q})$ then $-\omega^* \in \text{supp}(\mathbb{Q})$ and $\mathbb{Q}[\omega = \omega^*] = \mathbb{Q}[\omega = -\omega^*]$.*

Remark 2.4. *The properties of Brownian Motion imply that if $|\beta| = 2k + 1$ for some integer k , then $\mathbb{E}[J_{0,1}^\beta] = 0$. Cubature formulas with the symmetry properties will therefore approximate exactly iterated integrals of any odd degree.*

Lyons and Victoir [29], showed that the cubature method in the space $C^0([0, t], \mathbb{R}^d)$ provides a weak approximation to the Brownian motion with an error bounded by some power of the time length of approximation t . The construction can then be iteratively applied on small intervals to obtain an approximation with a good control for a time interval $[0, T]$ with arbitrary length. This motivates the following definition.

Definition 2.5 (Cubature measure). *Let $0 = t_0 < \dots < t_n = T$, and let $h_i = t_{i+1} - t_i$. Given a cubature formula $\tilde{\mathbb{Q}}$ represented by a set of weights $\theta_1, \dots, \theta_\kappa$ associated to a set of paths $\omega^1, \dots, \omega^\kappa$ in $C_{\text{bv}}^0([0, 1], \mathbb{R}^d)$, we build the probability space $(\tilde{\Omega}, \tilde{\mathbb{Q}})$, where $\tilde{\Omega} = C_{\text{bv}}^0([0, T], \mathbb{R}^d)$ and $\tilde{\mathbb{Q}}$ is a finite measure with support on paths indexed by $\eta \in \{1, \dots, \kappa\}^n$ given by*

$$\hat{\omega}^\eta(t) = \sum_{i=1}^n h_i^{1/2} \omega^{\eta_i} \left(\frac{(t_{i+1} \vee (t \wedge t_i)) - t_i}{h_i} \right) \mathbf{1}_{\{t > t_{i-1}\}}$$

with associated probability $\hat{\theta}_\eta = \theta_{\eta_1} \theta_{\eta_2} \dots \theta_{\eta_n}$.

Let us emphasize that the cubature measure constructed according to Definition 2.5 depends on the cubature formula $\tilde{\mathbb{Q}}$ and on the grid $\mathbb{T} = \{t_0, \dots, t_n\}$.

2.3. Forward Error expansion. Let us start by remarking that in all our development K, K', \dots denote constants that might depend on the parameters of the problem (i.e. T, f, g, b, σ, x_0) or on cubature measure that we assume fixed, but not on the parameters of the scheme. We use the convention that their value might change from line to line.

Proposition 2.6 (Regularity). *Under Assumption 1.1, there exists a constant K such that for all for all $(t, x) \in [0, T] \times \mathbb{R}^d$ and $0 < k \leq M$,*

$$(22) \quad \left\| \mathbb{E} \left[g(X_T^{t,x}) \right] \right\|_{k,\infty} \leq K \|g(x)\|_{k,\infty}.$$

Under Assumption 1.2, there exists a constant K such that for all for all $(t, x) \in [0, T] \times \mathbb{R}^d$ and $0 < k \leq M$,

$$(23) \quad \left\| \mathbb{E} \left[g(X_T^{t,x}) \right] \right\|_{k,\infty} \leq K |g|_{\text{Lip}}(T - t)^{-\frac{k-1}{2}}.$$

If in addition g has bounded derivatives up to order p , then

$$(24) \quad \left\| \mathbb{E} \left[g(X_T^{t,x}) \right] \right\|_{k,\infty} \leq K |\nabla^p g|_\infty (T - t)^{-\frac{k-p}{2}}.$$

Equation (22) is a consequence of the iterated application of Ito's formula, under Assumption 1.1. The claims under the UFG condition are proved in [18]: (23) is proved in Corollary 78 and (24) is deduced from extending the arguments of Corollary 32 with Corollary 78. The reader may refer to the PhD thesis [30] for further results on gradient bounds under alternative conditions.

2.3.1. *One-step expansion.* The following approximation result is a restatement of the results in [29] (see also Section 3.4 in [18]).

Proposition 2.7. *Under Assumption 1.1 (resp. 1.2), let $\psi : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ be a bounded function in \bar{C}_b^{m+2} uniformly in t . Then, for any $i \leq n-1$ (resp. $i \leq n-2$),*

$$(25) \quad \begin{aligned} & |Q_{t_i, t_{i+1}} \psi(t_i, x) - P_{t_i, t_{i+1}} \psi(t_i, x)| \\ & \leq K_{m+1} h_i^{\frac{m+1}{2}} \|\psi\|_{i; m+1, \infty} + K_{m+2} h_i^{\frac{m+2}{2}} \|\psi\|_{i; m+2, \infty} \end{aligned}$$

where $\|\cdot\|_{i, \infty}$ is defined in (21) and $K_j = \sum_{\beta \in \mathcal{A}_j \setminus \mathcal{A}_{j-1}} |C_\beta|$, with $C_\beta = \mathbb{E}^{\mathbb{Q}} I_{0,1}^\beta - \mathbb{E} J_{0,1}^\beta$.

If in addition $\psi \in \bar{C}_b^{m+3}$,

$$(26) \quad \begin{aligned} & \left| Q_{t_i, t_{i+1}} \psi(t_i, x) - P_{t_i, t_{i+1}} \psi(t_i, x) - h_i^{\frac{m+1}{2}} \sum_{\beta \in \mathcal{A}_{m+1} \setminus \mathcal{A}_m} C_\beta V^\beta \psi(t_i, x) \right| \\ & \leq K_{m+2} h_i^{\frac{m+2}{2}} \|\psi\|_{i; m+2, \infty} + K_{m+3} h_i^{\frac{m+3}{2}} \|\psi\|_{i; m+3, \infty} \end{aligned}$$

Moreover, if m is odd, the cubature measure \mathbb{Q} is symmetric and $\psi \in \bar{C}_b^{m+4}$, then

$$(27) \quad \begin{aligned} & \left| Q_{t_i, t_{i+1}} \psi(t_i, x) - P_{t_i, t_{i+1}} \psi(t_i, x) - h_i^{\frac{m+1}{2}} \sum_{\beta \in \mathcal{A}_{m+1} \setminus \mathcal{A}_m} C_\beta V^\beta \psi(t_i, x) \right| \\ & \leq K_{m+3} h_i^{\frac{m+3}{2}} \|\psi\|_{i; m+3, \infty} + K_{m+4} h_i^{\frac{m+4}{2}} \|\psi\|_{i; m+4, \infty}. \end{aligned}$$

Proof. From the Taylor-Stratonovich expansion (see Theorem 5.6.1 in [27]) applied to ψ , one gets, for any $x \in \mathbb{R}^d$,

$$P_{t_i, t_{i+1}} \psi(t_i, x) = \mathbb{E} \left[\psi(t_{i+1}, X_{t_{i+1}}^{t_i, x}) \right] = \sum_{\beta \in \mathcal{A}_m} V^\beta \psi(t_i, x) \mathbb{E} J_{t_i, t_{i+1}}^\beta + \sum_{\beta \in \partial \mathcal{A}_m} \mathbb{E} J_{t_i, t_{i+1}}^\beta (V^\beta \psi(\cdot, X_{t_i}^{t_i, x})).$$

Similarly, a Taylor expansion and the definition of \hat{X} in (5) shows

$$Q_{t_i, t_{i+1}} \psi(t_i, x) = \mathbb{E}^{\mathbb{Q}} \left[\psi(t_{i+1}, \hat{X}_{t_{i+1}}^{t_i, x}) \right] = \sum_{\beta \in \mathcal{A}_m} V^\beta \psi(t_i, x) \mathbb{E}^{\mathbb{Q}} I_{t_i, t_{i+1}}^\beta + \sum_{\beta \in \partial \mathcal{A}_m} \mathbb{E}^{\mathbb{Q}} I_{t_i, t_{i+1}}^\beta (V^\beta \psi(\cdot, \hat{X}_{t_i}^{t_i, x})).$$

From the construction of the cubature measure in Definition 2.5 and the definition of the cubature formula, we get that $\mathbb{E}^{\mathbb{Q}} I_{t_i, t_{i+1}}^\beta = \mathbb{E} J_{t_i, t_{i+1}}^\beta$ for all $\beta \in \mathcal{A}_m$. Hence, by using the regularity assumptions on ψ , the scaling properties of the cubature and the Markov property.

$$\begin{aligned}
(28) \quad & [Q_{t_i, t_{i+1}} - P_{t_i, t_{i+1}}] \psi(t_i, x) \\
&= \sum_{\beta \in \partial \mathcal{A}_m} \mathbb{E}^{\hat{Q}} I_{t_i, t_{i+1}}^\beta (V^\beta \psi(\cdot, \hat{X})) - \mathbb{E} J_{t_i, t_{i+1}}^\beta (V^\beta \psi(\cdot, X^{t_i, x})) \\
&\leq \sum_{\beta \in \partial \mathcal{A}_m} h_i^{\frac{\|\beta\|}{2}} \sup_{t \in [t_i, t_{i+1})} |V^\beta \psi(t, \cdot)|_\infty \left(\mathbb{E}^{\hat{Q}} I_{0,1}^\beta - \mathbb{E} J_{0,1}^\beta \right) \\
&\leq K_{m+1} h_i^{\frac{m+1}{2}} \|\psi\|_{i; m+1, \infty} + K_{m+2} h_i^{\frac{m+2}{2}} \|\psi\|_{i; m+2, \infty}
\end{aligned}$$

where for $j = m+1, m+2$,

$$K_j = \sum_{\beta \in \mathcal{A}_j \setminus \mathcal{A}_{j-1}} \left| \mathbb{E}^{\hat{Q}} I_{0,1}^\beta - \mathbb{E} J_{0,1}^\beta \right|.$$

Having shown (25), let us now assume $\psi \in \bar{C}_b^{m+3}$. Then we can expand the terms in \mathcal{A}_{m+1} in the equality in (28), and apply the scaling properties of the cubature to get

$$\begin{aligned}
& \left| [Q_{t_i, t_{i+1}} - P_{t_i, t_{i+1}}] \psi(t_i, x) - \sum_{\beta \in \mathcal{A}_{m+1} \setminus \mathcal{A}_m} h_i^{\frac{\|\beta\|}{2}} V^\beta \psi(t_i, x) \left\{ \mathbb{E}^{\hat{Q}} I_1^\beta - \mathbb{E} J_1^\beta \right\} \right| \\
&= \sum_{\beta \in \partial \mathcal{A}_{m+1}} \mathbb{E}^{\hat{Q}} I_{t_i, t_{i+1}}^\beta (V^\beta \psi(\cdot, \hat{X})) - \mathbb{E} J_{t_i, t_{i+1}}^\beta (V^\beta \psi(\cdot, X^{t_i, x})) \\
&\leq K_{m+2} h_i^{\frac{m+2}{2}} \|\psi\|_{i; m+2, \infty} + K_{m+3} h_i^{\frac{m+3}{2}} \|\psi\|_{i; m+3, \infty}.
\end{aligned}$$

Finally, if ψ is $m+4$ times differentiable, then we can expand once more and use the symmetry of the cubature formula and Remark 2.4 to obtain the result.

□

2.3.2. Global expansion: Proof of Theorem 1.3. Using the previous one-step results, we can study an explicit error expansion for the error in the cubature approximation of X for several steps, thus extending [33] to the cubature approximation. We analyze this property under both Assumption 1.1 and 1.2. We start with the following control result.

Lemma 2.8. *Suppose that either Assumption 1.1 or Assumption 1.2 hold. Assume that (t_0, \dots, t_n) are defined as in (7), with $\gamma = 1$ if Assumption 1.1 holds and $\gamma \geq m$ otherwise. Then, there is a constant K such that, for all $0 < k \leq M$,*

$$\sum_{i=0}^{n-2} h_i^{\frac{k}{2}} \|P_{\cdot, t_n} g\|_{i; k, \infty} \leq K n^{-\frac{k-1}{2}}$$

Proof. When Assumption 1.1 holds, we have that $h_i = T n^{-1}$ and from Proposition 2.6,

$$\sum_{i=0}^{n-2} h_i^{\frac{k}{2}} \|P_{\cdot, t_n} g\|_{i; k, \infty} \leq \sum_{i=0}^{n-2} K n^{-\frac{k}{2}} \|g(\cdot)\|_{k, \infty} \leq K' n^{-\frac{k-1}{2}}.$$

On the other hand, if Assumption 1.2 holds, we have from Proposition 2.6 and Corollary A.3,

$$\sum_{i=0}^{n-2} h_i^{\frac{k}{2}} \|P_{\cdot, t_n} g\|_{i; k, \infty} \leq K' n^{-\frac{k-1}{2}}.$$

□

Before proving the main expansion result for this part i.e. Theorem 1.3, let us make precise the shape of the leading coefficient Ψ^{lin} in the expansion.

Definition 2.9.

i. Under Assumption 1.1,

$$(29) \quad \Psi_T^{lin}(s, x) := T^{\frac{m-1}{2}} \mathbb{E} \left[\sum_{\beta \in \mathcal{A}_{m+1} \setminus \mathcal{A}_m} \int_s^T C_\beta V^\beta [P_{t,T} g(X_t^{0,x})] dt \right].$$

ii. Under Assumption 1.2

$$\Psi_T^{lin}(s, x) := (T^{\frac{1}{\gamma}} \gamma)^{\frac{m-1}{2}} \mathbb{E} \left[\sum_{\beta \in \mathcal{A}_{m+1} \setminus \mathcal{A}_m} \int_s^T C_\beta V^\beta [P_{t,T} g(X_t^{0,x})] (T-t)^{(1-\frac{1}{\gamma})\frac{m-1}{2}} dt \right].$$

We are ready to prove the global error expansion result for the conditional expectation using the cubature method.

Proof of Theorem 1.3 To prove the first statement, let \mathcal{L} denote the infinitesimal operator associated with the diffusion X .

Let u be given as the solution of the linear equation $\partial_t u(t, x) + \mathcal{L}u(t, x) = 0$ defined on $[0, T)$ with boundary condition $u(T, x) = g(x)$. Clearly, for any $i = 0, \dots, n-1$, we have $P_{t_i, t_{i+1}} u(t_i, x) = u(t_i, x)$. In particular,

$$u(0, x) = P_{0,T} u(0, x) = \mathbb{E}[u(T, X_T^{0,x})] = \mathbb{E}[g(X_T^{0,x})] = P_{0,T} g(x)$$

Note that given that $M \geq m+2$, Assumptions 1.1 (resp. 1.2) and Proposition 2.6 imply that $u(t_i, \cdot)$ has bounded derivatives of all order up to $m+2$, for $i < n$.

Using the definition of the function u and the properties of the family of operators Q and P , we can then reformulate the error term as a telescopic sum,

$$(30) \quad \begin{aligned} \mathbb{E}^{\hat{\mathbb{Q}}}[g(\hat{X}_T^{0,x})] - \mathbb{E}[g(X_T^{0,x})] &= Q_{0,T} u(0, x) - P_{0,T} u(0, x) \\ &= \sum_{i=0}^{n-1} Q_{0,t_i} \{ [Q_{t_i, t_{i+1}} - P_{t_i, t_{i+1}}] P_{t_{i+1}, T} u \}(0, x) \\ &= \sum_{i=0}^{n-2} Q_{0,t_i} \{ [Q_{t_i, t_{i+1}} - P_{t_i, t_{i+1}}] u \}(0, x) + Q_{0,t_{n-1}} \{ [Q_{t_{n-1}, t_n} - P_{t_{n-1}, t_n}] g \}(x). \end{aligned}$$

The last term in 30 is simply an average of a one step cubature approximation. Under Assumption (1.1), g is regular enough to use the one-step cubature approximation result in Proposition (2.7), hence

$$(31) \quad |[Q_{t_{n-1}, t_n} - P_{t_{n-1}, t_n}] g(x)|_\infty \leq K h_{n-1}^{\frac{m+1}{2}} \|g\|_{m+2, \infty} = K n^{-\frac{m+1}{2}} \|g\|_{m+2, \infty}.$$

The same result cannot be applied under Assumption (1.2) though, as g is not supposed to be regular. However using the Lipschitz regularity of g , the bounded variation of the cubature formula and its rescaling, we get

$$(32) \quad \begin{aligned} |[Q_{t_{n-1}, t_n} - P_{t_{n-1}, t_n}] g(x)|_\infty &\leq |Q_{t_{n-1}, t_n} [g](x) - g(x)|_\infty + |P_{t_{n-1}, t_n} [g](x) - g(x)|_\infty \\ &\leq K |g|_{Lip} h_{n-1}^{\frac{1}{2}} = K |g|_{Lip} n^{-\frac{\gamma}{2}} \end{aligned}$$

To treat the first term in the last term in 30, (25) and Lemma 2.8, we have that

$$(33) \quad \begin{aligned} \sum_{i=0}^{n-2} Q_{0,t_i} \{ [Q_{t_i, t_{i+1}} - P_{t_i, t_{i+1}}] u \}(0, x) &\leq K_{m+1} \sum_{i=0}^{n-1} h_i^{\frac{m+1}{2}} \|u\|_{i; m+1, \infty} + K_{m+2} \sum_{i=0}^{n-1} h_i^{\frac{m+2}{2}} \|u\|_{i; m+2, \infty} \\ &\leq K' n^{-\frac{m-1}{2}}. \end{aligned}$$

The first claim is thus proven.

For the second claim, we use the extra regularity assumptions to apply successively, Jensen's inequality, properties of expectation operators, the claim on symmetric cubatures of Proposition 2.7 and Lemma 2.8, to write

$$\begin{aligned}
(34) \quad & \left| \sum_{i=0}^{n-2} Q_{0,t_i} \{ [Q_{t_i,t_{i+1}} - P_{t_i,t_{i+1}}] u \} (0, x) - \sum_{i=0}^{n-2} \sum_{\beta \in \mathcal{A}_{m+1} \setminus \mathcal{A}_m} C_\beta h_i^{\frac{m+1}{2}} Q_{0,t_i} [V^\beta u] (0, x) \right| \\
& \leq \sum_{i=0}^{n-2} Q_{0,t_i} \left| \{ Q_{t_i,t_{i+1}} - P_{t_i,t_{i+1}} \} u - \sum_{\beta \in \mathcal{A}_{m+1} \setminus \mathcal{A}_m} C_\beta h_i^{\frac{m+1}{2}} [V^\beta u] \right| (0, x) \\
& \leq \sum_{i=0}^{n-2} \left| \{ Q_{t_i,t_{i+1}} - P_{t_i,t_{i+1}} \} u - \sum_{\beta \in \mathcal{A}_{m+1} \setminus \mathcal{A}_m} C_\beta h_i^{\frac{m+1}{2}} [V^\beta u] (0, x) \right| \\
& \leq K_{m+3} \sum_{i=0}^{n-2} h_i^{\frac{m+3}{2}} \|u\|_{i;m+3,\infty} + K_{m+4} \sum_{i=0}^{n-2} h_i^{\frac{m+4}{2}} \|u\|_{i;m+4,\infty} \\
& \leq K n^{-\frac{m+1}{2}}.
\end{aligned}$$

Now, we are going to show that

$$(35) \quad \left| \sum_{i=0}^{n-2} \sum_{\beta \in \mathcal{A}_{m+1} \setminus \mathcal{A}_m} C_\beta h_i^{\frac{m+1}{2}} \left[Q_{0,t_i} [V^\beta u] (0, x) - P_{0,t_i} [V^\beta u] (0, x) \right] \right| \leq K n^{-(m-1)}.$$

To prove this, note first that from the regularity of $V^\beta u$, the definition of the seminorm and the bounds in Proposition 2.6 we have

$$\|V^\beta u(t, \cdot)\|_{m+1,\infty} = \|V^\beta P_{t,Tg}(\cdot)\|_{m+1,\infty} \leq K \|P_{t,Tg}(\cdot)\|_{m+1+\|\beta\|,\infty} = K \|u(t, \cdot)\|_{m+1+\|\beta\|,\infty}.$$

Hence, using equation (33) and reordering, we have

$$\begin{aligned}
(36) \quad & \left| \sum_{i=0}^{n-2} \sum_{\beta \in \mathcal{A}_{m+1} \setminus \mathcal{A}_m} C_\beta h_i^{\frac{m+1}{2}} \left[Q_{0,t_i} [V^\beta u] (0, x) - P_{0,t_i} [V^\beta u] (0, x) \right] \right| \\
& \leq K \sum_{i=0}^{n-2} h_i^{\frac{m+1}{2}} \sum_{\beta \in \mathcal{A}_{m+1} \setminus \mathcal{A}_m} \left[\sum_{j=0}^{i-1} h_j^{\frac{m+1}{2}} \|V^\beta u\|_{j;m+1,\infty} + \sum_{j=0}^{i-1} h_j^{\frac{m+2}{2}} \|V^\beta u\|_{j;m+2,\infty} \right] \\
& \leq K \sum_{i=0}^{n-2} h_i^{\frac{m+1}{2}} \left[\sum_{j=0}^{i-1} h_j^{\frac{m+1}{2}} \|u\|_{j;2m+2,\infty} + \sum_{j=0}^{i-1} h_j^{\frac{m+2}{2}} \|u\|_{j;2m+3,\infty} \right] \\
& \leq K \sum_{j=0}^{n-2} \left(\sum_{i=j+1}^{n-2} h_i^{\frac{m+1}{2}} \right) h_j^{\frac{m+1}{2}} \|u\|_{j;2m+2,\infty} + \sum_{j=0}^{i-1} \left(\sum_{i=j+1}^{n-2} h_i^{\frac{m+1}{2}} \right) h_j^{\frac{m+2}{2}} \|u\|_{j;2m+3,\infty}.
\end{aligned}$$

If Assumption 1.1 holds, we deduce (35) directly using the boundedness in Proposition 2.6. Under Assumption 1.2, we deduce from (59) in Lemma A.4

$$\begin{aligned}
(37) \quad & \sum_{j=0}^{n-2} \left(\sum_{i=j+1}^{n-1} h_i^{\frac{m+1}{2}} \right) h_j^{\frac{m+1}{2}} \|u(t_j, \cdot)\|_{2m+2, \infty} \\
& \leq K \left(\frac{T^{1/\gamma} \gamma}{n} \right)^{\frac{m-1}{2}} \sum_{j=0}^{n-2} \left(\int_{t_{j+1}}^T (T-t)^{\frac{m-1}{2}(1-\frac{1}{\gamma})} dt \right) h_j^{\frac{m+1}{2}} (T-t_j)^{-\frac{2m+1}{2}} \\
& \leq K \left(\frac{T^{1/\gamma} \gamma}{n} \right)^{\frac{m-1}{2}} \sum_{j=0}^{n-2} (T-t_{j+1})^{-\frac{m-1}{2}(1-\frac{1}{\gamma})+1} h_j^{\frac{m+1}{2}} (T-t_j)^{-\frac{2m+1}{2}} \\
& \leq K n^{-(m-1)},
\end{aligned}$$

where the last inequality follows from (the proof of) Corollary A.3. A similar development can be applied to the second term in the last line of (36), to deduce (35) in this case.

Finally, the claim follows from (34), (35), (31) (or (32) under Assumption 1.2) and

$$\left| n^{-\frac{m-1}{2}} \Psi_T^{lin}(0, x) - \sum_{i=0}^{n-2} \sum_{\beta \in \mathcal{A}_{m+1} \setminus \mathcal{A}_m} C_\beta h_i^{\frac{m+1}{2}} P_{0, t_i}[V^\beta u](0, x) \right| \leq K n^{-\frac{m+1}{2}}$$

To conclude, we can apply Lemma A.4 with $\psi(t) = C_\beta V^\beta [P_{t, T} g(X_t^{0, x})]$: the control on $|\psi|$ readily follows from Proposition 2.6. Moreover, since for each $\beta \in \mathcal{A}_{M-2}$, we have

$$\partial_t E[V^\beta u(t, X_t)] = \mathbb{E}[L^0 V^\beta u(t, X_t)] = \mathbb{E} \left[V^{0 \star \beta} u(t, X_t) + \sum_{i=1}^d V^{(i, i) \star \beta} u(t, X_t) \right],$$

it follows that ψ has well defined locally bounded first order derivatives in $[0, T]$, and hence it is of bounded variation in $[0, T - \epsilon]$ for all $\epsilon > 0$ as required. \square

3. STUDY OF THE BACKWARD APPROXIMATION

Our goal in this section is to study the error terms in the approximation of the backward function, given by

$$\Delta \hat{u}_i(x) := \hat{u}_i(x) - u(t_i, x).$$

An inspection of the proof of Theorem 1.3 shows that we used the linearity properties of the operators P and Q to decompose the global error in a sum of one-step errors.

We follow a similar idea to expand the error in the case of backward equations. However, the non-linearity will change the type of decomposition in terms of one-step errors that can be achieved, and will require some additional approximations. For this purpose, let us define

$$\tilde{v}_i(x) = \mathbb{E}^{\hat{\mathbb{Q}}} \left[u(t_{i+1}, \hat{X}_{t_{i+1}}^{t_i, x}) \frac{\Delta \hat{\omega}_i}{h_i} \right],$$

and set

$$\Delta \tilde{v}_i(x) := \hat{v}_i(x) - \tilde{v}_i(x) = \frac{1}{h_i} \mathbb{E}^{\hat{\mathbb{Q}}} \left[\Delta \hat{u}_{i+1}(\hat{X}_{t_{i+1}}^{t_i, x}) \Delta \hat{\omega}_i \right].$$

Then, we can re-write the decoupling function evaluated at the grid times as the solution of a perturbed scheme, namely

$$u(t_i, x) = \mathbb{E}^{\hat{\mathbb{Q}}} \left[u(t_{i+1}, \hat{X}_{t_{i+1}}^{t_i, x}) + h_i f(x, u(t_i, x), \tilde{v}_i(x)) \right] + \zeta_i(x)$$

with

$$\begin{aligned} \zeta_i(x) := & \mathbb{E} \left[u(t_{i+1}, X_{t_{i+1}}^{t_i, x}) \right] - \mathbb{E}^{\hat{\mathbb{Q}}} \left[u(t_{i+1}, \hat{X}_{t_{i+1}}^{t_i, x}) \right] + h_{i+1} \{ f(x, u(t_i, x), \check{v}_i(x)) - f(x, u(t_i, x), \tilde{v}_i(x)) \} \\ & + \mathbb{E} \left[\int_{t_i}^{t_{i+1}} \{ f(X_s^{t_i, x}, u(s, X_s^{t_i, x}), v(s, X_s^{t_i, x})) - f(x, u(t_i, x), \check{v}_i(x)) \} ds \right] \end{aligned}$$

$$\text{where } \check{v}_i(x) = \mathbb{E} \left[\frac{\Delta W_i}{h_i} u(t_{i+1}, X_{t_{i+1}}^{t_i, x}) \right].$$

Let us recall the following regularity result valid under our set of assumptions.

Proposition 3.1 (Space regularity of u). *Suppose either Assumption 1.1 or 1.2 hold. For all $0 < k \leq M$, and $\alpha \in \mathcal{A}_k$, the function $V_\alpha u$ is well-defined. Moreover,*

- Under Assumption 1.1, there exists a constant C such that for all $t < T$

$$\|u(t, \cdot)\|_{k, \infty} \leq C \|g\|_{k, \infty}.$$

- Under Assumption 1.2, there exists a constant C such that for all $t < T$

$$\|u(t, \cdot)\|_{k, \infty} \leq C \|g\|_{\text{Lip}} (T - t)^{-\frac{k-1}{2}}.$$

The proposition under Assumption 1.1 is proved by repeated application of the results in [32] (in particular Theorem 3.2. and the BSDE representation for the derivative of u). The proposition under Assumption 1.2 is proved in Theorem 1.4 [14].

Lemma 3.2 (One step controls). *Let $\hat{\mathbb{Q}}$ be a cubature measure constructed from a symmetric cubature formula of order $m \geq 3$. Let $i \leq n - 2$. Under Assumptions 1.1 or 1.2 with $M \geq 7$ we have*

$$\begin{aligned} i. & \mathbb{E} \left[u(t_{i+1}, X_{t_{i+1}}^{t_i, x}) \right] - \mathbb{E}^{\hat{\mathbb{Q}}} \left[u(t_{i+1}, \hat{X}_{t_{i+1}}^{t_i, x}) \right] = \varphi^e(t_i, x) h_i^2 + \xi^e(t_i, x) h_i^3, \\ ii. & f(t_i, x, u(t_i, x), \check{v}_i(x)) - f(t_i, x, u(t_i, x), \tilde{v}_i(x)) = \varphi^f(t_i, x) h_i + \xi^f(t_i, x) h_i^2, \\ iii. & \mathbb{E} \left[\int_{t_i}^{t_{i+1}} \{ f(s, X_s^{t_i, x}, u(s, X_s^{t_i, x}), v(s, X_s^{t_i, x})) - f(t_i, x, u(t_i, x), \check{v}_i(x)) \} ds \right] = \varphi^\tau(t_i, x) h_i^2 + \xi^\tau(t_i, x) h_i^3, \end{aligned}$$

where ξ^e, ξ^f, ξ^τ are bounded measurable functions,

$$\varphi^e(t_i, x) = - \sum_{\beta \in \mathcal{A}_4 \setminus \mathcal{A}_3} C_\beta V^\beta u(t_i, x)$$

$$\varphi^f(t_i, x) = \sum_{j=1}^d \partial_{z_j} f(t_i, x, u(t_i, x), \check{v}_i(x)) \left\{ \sum_{\beta \in \mathcal{A}_3 \setminus \mathcal{A}_2} \left(C_{j \star \beta} + \sum_{k=1}^{|\beta|} C_{\beta_{1:k} \star j \star \beta_{k+1:|\beta|_0}} \right) V^\beta u(t_i, x) \right\}$$

$$\begin{aligned} \varphi^\tau(t_i, x) = & \sum_{\beta \in \mathcal{A}_2 \setminus \mathcal{A}_1} \left(V^{\beta \star (0)} + \sum_{j=1}^d V^{\beta \star (j, j)} \right) u(t_i, x) \\ & + \sum_{j=1}^d \sum_{\beta \in \mathcal{A}_3 \setminus \mathcal{A}_2} \partial_{z_j} f(t_i, x, u(t_i, x), v(t_i, x)) \left(V^{j \star \beta} + \sum_{k=1}^{|\beta|} V^{\beta_{1:k} \star j \star \beta_{k+1:|\beta|_0}} \right) u(t, x) \end{aligned}$$

$$v(t, x) := \sum_{k=1}^d V^{(k)} u(t, x);$$

the constants $\{C_\beta\}_{\beta \in \mathcal{A}_m}$ are defined as in Proposition 2.7. Moreover, there exist constants K, K', K_6, K_7 such that

$$\begin{aligned} |\xi^e|_\infty &\leq K_6 \|u\|_{i;6,\infty} + K_7 h_i^{1/2} \|u\|_{i;7,\infty}, \\ |\xi^f|_\infty &\leq K \|u\|_{i;5,\infty}, \\ |\xi^\tau|_\infty &\leq K' (\|u\|_{i;5,\infty} + \|u\|_{i;6,\infty}). \end{aligned}$$

Remark 3.3. Under Assumption 1.1, the claim also holds for $i = n - 1$. Under Assumption 1.2, ii. and iii. still hold, while using Lipschitz continuity of g , we can replace i. by

$$i'. \quad \left| \mathbb{E} \left[u(t_{i+1}, X_{t_{i+1}}^{t_i, x}) \right] - \mathbb{E}^\mathbb{Q} \left[u(t_{i+1}, \hat{X}_{t_{i+1}}^{t_i, x}) \right] \right| \leq K n^{-\frac{\gamma}{2}} |g|_{Lip}.$$

Proof of Lemma 3.2.

- i. The claim is a consequence of the cubature control in Proposition 2.7, and the regularity of u stated in Proposition 3.1.
- ii. The integration by parts properties of the Stratonovich integral imply that

$$\Delta W_i^j J_{t_i, t_{i+1}}^\beta = J_{t_i, t_{i+1}}^j J_{t_i, t_{i+1}}^\beta = J_{t_i, t_{i+1}}^{j \star \beta} + \sum_{k=1}^{|\beta|} J_{t_i, t_{i+1}}^{\beta_{1:k} \star j \star \beta_{k+1:|\beta|_0}}.$$

Hence, by definition of the symmetric cubature measure and using a Stratonovich Taylor expansion as in the proof of Proposition 2.7, we have that

$$\begin{aligned} (38) \quad \check{v}_i^j(x) - \tilde{v}_i^j(x) &= \mathbb{E} \left[\frac{\Delta W_i^j}{h_i} u(t_{i+1}, X_{t_{i+1}}^{t_i, x}) \right] - \mathbb{E}^\mathbb{Q} \left[\frac{\Delta \hat{W}_i^j}{h_i} u(t_{i+1}, \hat{X}_{t_{i+1}}^{t_i, x}) \right] \\ &\leq h_i \sum_{\beta \in \mathcal{A}_3 \setminus \mathcal{A}_2} V^\beta u(t_i, x) \left(C_{j \star \beta} + \sum_{k=1}^{|\beta|} C_{\beta_{1:k} \star j \star \beta_{k+1:|\beta|_0}} \right) + K_6 h_i^2 \|u\|_{i;5,\infty}. \end{aligned}$$

Denoting

$$\check{\alpha}_i^{j,k}(x) = \int_0^1 (1 - \lambda) \partial_{z_j, z_k} f(t_i, x, u(t_i, x), \check{v}_i(x) + \lambda(\tilde{v}_i(x) - \check{v}_i(x))) d\lambda,$$

we have

$$\begin{aligned} f(t_i, x, u(t_i, x), \tilde{v}_i(x)) - f(t_i, x, u(t_i, x), \check{v}_i(x)) \\ = \sum_{j=1}^d \partial_{z_j} f(t_i, x, u(t_i, x), \check{v}_i(x)) (\tilde{v}_i^j(x) - \check{v}_i^j(x)) + \sum_{j=1}^d \sum_{k=1}^d \check{\alpha}_i^{j,k}(x) (\tilde{v}_i^k(x) - \check{v}_i^k(x)) (\tilde{v}_i^j(x) - \check{v}_i^j(x)) \end{aligned}$$

and replacing (38) this leads to

$$\begin{aligned} f(x, u(t_i, x), \tilde{v}_i(x)) - f(x, u(t_i, x), \check{v}_i(x)) \\ = h_i \sum_{j=1}^d \partial_{z_j} f(x, u(t_i, x), \check{v}_i(x)) \left\{ \sum_{\beta \in \mathcal{A}_3 \setminus \mathcal{A}_2} V^\beta u(t_i, x) \left(C_{j \star \beta} + \sum_{k=1}^{|\beta|} C_{\beta_{1:k} \star j \star \beta_{k+1:|\beta|_0}} \right) \right\} + \xi^f(t_i, x) h_i^2 \end{aligned}$$

where

$$|\xi^f|_\infty \leq K_6 \|u\|_{i;5,\infty}.$$

iii. Note that the regularity of u , implies that it solves classically the equation

$$f(t, x, u(t, x), v(t, x)) = L^{(0)}u(t, x) = \left(V^{(0)} + \sum_{j=1}^d V^{(j,j)} \right) u(t, x).$$

Hence, Fubini's theorem implies

$$\begin{aligned} & \mathbb{E} \left[\int_{t_i}^{t_{i+1}} [f(s, X_s^{t_i, x}, u(s, X_s^{t_i, x}), v(s, X_s^{t_i, x})) - f(t_i, x, u(t_i, x), v_i(x))] ds \right] \\ &= \int_{t_i}^{t_{i+1}} \mathbb{E} \left[\left(V^{(0)} + \sum_{j=1}^d V^{(j,j)} \right) u(s, X_s^{t_i, x}) - \left(V^{(0)} + \sum_{j=1}^d V^{(j,j)} \right) u(t_i, x) \right] ds \\ &= h_i^2 \sum_{\beta \in \mathcal{A}_2 \setminus \mathcal{A}_1} V^\beta \left(V^{(0)} + \sum_{j=1}^d V^{(j,j)} \right) u(t_i, x) + R_i(x) h_i^3. \end{aligned}$$

where $|R_i|_\infty < K \|u\|_{6, \infty}$. Moreover,

$$\begin{aligned} (39) \quad v^j(t_i, x) - \check{v}_i^j(x) &= \left\{ V^j u(t_i, x) - \mathbb{E} \left[\frac{\Delta W_i^j}{h_i} u(t_i, X_{t_{i+1}}^{t_i, x}) \right] \right\} \\ &= -h_i \sum_{\beta \in \mathcal{A}_3 \setminus \mathcal{A}_2} \left(V^{j \star \beta} + \sum_{k=1}^{|\beta|} V^{\beta_{1:k} \star j \star \beta_{k+1:|\beta|_0}} \right) u(t, x) + R_i^v(x) h_i^2 \end{aligned}$$

with

$$|R_i^v|_\infty \leq \#(\partial \mathcal{A}_3 \cap \mathcal{A}_5) \cdot \|u\|_{i;5, \infty}.$$

Hence, by a similar argument as before, we have

$$\begin{aligned} f(t_i, x, u(t_i, x), \check{v}_i(x)) &= f(t_i, x, u(t_i, x), v(t_i, x)) \\ &\quad + h_i \sum_{j=1}^d \sum_{\beta \in \mathcal{A}_3 \setminus \mathcal{A}_2} \partial_{z_j} f(t_i, x) \left(V^{j \star \beta} + \sum_{k=1}^{|\beta|} V^{\beta_{1:k} \star j \star \beta_{k+1:|\beta|_0}} \right) u(t, x) \\ &\quad + R'_i(x) h_i^2. \end{aligned}$$

where we wrote $\partial_{z_j} f(t_i, x)$ to mean the quantity $\partial_{z_j} f(t_i, x, u(t_i, x), v(t_i, x))$, and with $|R'_i|_\infty < \|\partial_{z_j} f\|_\infty |R_i^v|_\infty \|u\|_{i;5, \infty}$. The claim then follows. \square

With the notation of the previous lemma, we define

$$(40) \quad \varphi(t, x) = \varphi^e(t, x) + \varphi^f(t, x) + \varphi^\tau(t, x),$$

$$(41) \quad \xi(t, x) = \xi^e(t, x) + \xi^f(t, x) + \xi^\tau(t, x),$$

and $\zeta_i(x) = \varphi(t_i, x) h_i^2 + \xi(t_i, x) h_i^3$. Note that, if M is large enough

$$(42) \quad \|\varphi(t_i, \cdot)\|_{l, \infty} \leq K \|u\|_{i; l+4, \infty},$$

and

$$(43) \quad \|\xi(t_i, \cdot)\|_{l, \infty} \leq K \|u\|_{i; l+6, \infty}.$$

3.1. Error expansion. We now proceed to show the full error expansion under our set of assumptions. Let us recall the following result on the rate of convergence of the scheme (see [16] or [13]).

Lemma 3.4. *Let $\hat{\mathbb{Q}}$ be a cubature measure constructed from a symmetric cubature formula of order $m \geq 3$. Under Assumptions 1.1 or 1.2 with $M \geq 6$, (and $\gamma \geq 2$ in the case of Assumption 1.2), we have*

$$|\Delta \hat{u}_i|_\infty \leq K \left(|\Delta \hat{u}_{n-1}|_\infty + \sum_{j=i}^{n-2} \|u(t_j, \cdot)\|_{4,\infty} h_j^2 + \|u(t_j, \cdot)\|_{6,\infty} h_j^3 \right) \leq K' n^{-1}.$$

Moreover,

$$|\Delta \hat{u}_i|_\infty^2 + \sum_{j=i}^{n-1} h_j |\Delta \tilde{v}_j|_\infty^2 \leq K \left(|\Delta \hat{u}_{n-1}|_\infty^2 + \sum_{j=i}^{n-1} \|u(t_j, \cdot)\|_{4,\infty} h_j^2 + \|u(t_j, \cdot)\|_{6,\infty} h_j^3 \right)^2 \leq K' n^{-2}.$$

3.1.1. One-step expansion. We first prove the following key one-step expansion.

Lemma 3.5. *Under the assumptions of Lemma 3.2, we have that*

$$(44) \quad \Delta \hat{u}_i(x) = \mathbb{E}^{\hat{\mathbb{Q}}} \left[\hat{\Lambda}_{i,i+1}^x \Delta \hat{u}_{i+1}(\hat{X}_{t_{i+1}}^{t_i, x}) \right] - e^{h_i \eta^y(t_i, x)} \varphi(t_i, x) h_i^2 - e^{h_i \eta^y(t_i, x)} \xi(t_i, x) h_i^3 \\ - e^{h_i \eta^y(t_i, x)} R_i^{\text{exp}}(x) \Delta \hat{u}_i(x) + h_i e^{h_i \eta^y(t_i, x)} R_i^{u, \tilde{v}}(x),$$

where

$$(45) \quad \eta^y(t_i, x) := \partial_y f(t_i, x, u(t_i, x), \tilde{v}_i(x)), \quad \eta^z(t_i, x) := \partial_z f(t_i, x, u(t_i, x), \tilde{v}_i(x)), \\ \hat{\Lambda}_{i,j}^x := \prod_{k=i}^{j-1} \left[e^{h_k \eta^y(t_k, \hat{X}_{t_k}^{t_i, x})} (1 + \eta^z(t_k, \hat{X}_{t_k}^{t_i, x}) \Delta \hat{\omega}_k) \right], \text{ for } j > i,$$

and $\hat{\Lambda}_{i,i}^x := 1$. Moreover, for some constants $C^z, C^{\text{exp}}, C^u, C^{\tilde{v}}$

$$|R_i^z|_\infty \leq C^z |\Delta \hat{u}_{i+1}|_\infty h_i^2, \quad |R_i^{\text{exp}}|_\infty \leq C^{\text{exp}} h_i^2, \\ \text{and } |R_i^{u, \tilde{v}}|_\infty \leq C^u |\Delta \hat{u}_i|_\infty^2 + C^{\tilde{v}} |\Delta \tilde{v}_i|_\infty^2.$$

Proof. Since the exact solution satisfies a perturbed version of the scheme, we have

$$\Delta \hat{u}_i(x) = \mathbb{E}^{\hat{\mathbb{Q}}} \left[\hat{u}_i(\hat{X}_{t_{i+1}}^{t_i, x}) - u(t_{i+1}, \hat{X}_{t_{i+1}}^{t_i, x}) \right] - \varphi(t_i, x) h_i^2 - \xi(t_i, x) h_i^3 \\ + h_i [f(t_i, x, \hat{u}_i(x), \hat{v}_i(x)) - f(t_i, x, u(t_i, x), \tilde{v}_i(x))].$$

We compute, using the mean value theorem,

$$f(t_i, x, \hat{u}_i(x), \hat{v}_i(x)) - f(t_i, x, u(t_i, x), \tilde{v}_i(x)) = \eta^y(t_i, x) \Delta \hat{u}_i(x) + \eta^z(t_i, x) \Delta \tilde{v}_i(x) + R_i^{u, \tilde{v}}(x)$$

where

$$R_i^{u, \tilde{v}}(x) = \alpha_i^{yz} \Delta u_i(x) \Delta \tilde{v}_i(x) + \frac{1}{2} \alpha_i^{yy} |\Delta \hat{u}_i(x)|^2 + \frac{1}{2} \alpha_i^{zz} |\Delta \tilde{v}_i(x)|^2 \\ \leq C^u |\Delta \hat{u}_i(x)|^2 + C^{\tilde{v}} |\Delta \tilde{v}_i(x)|^2$$

for some $\alpha_i^{yz}, \alpha_i^{xy}, \alpha_i^{zz}$ with norm bounded by $\sup_{|\beta|_1 \leq 2} |D^\beta f|$.

This leads to

$$(46) \quad \Delta \hat{u}_i(x) (1 - h_i \eta^y(t_i, x)) = \mathbb{E}^{\hat{\mathbb{Q}}} \left[\Delta \hat{u}_i(\hat{X}_{t_{i+1}}^{t_i, x}) (1 + \Delta \hat{\omega}_i \eta^z(t_i, x)) \right] - \varphi(t_i, x) h_i^2 \\ - \xi(t_i, x) h_i^3 + h_i R_i^{u, \tilde{v}}(x).$$

On the other hand, notice that for h_i small enough,

$$(47) \quad \frac{1}{1 - h_i \eta^y(t_i, x)} = e^{h_i \eta^y(t_i, x)} + R_i^{\text{exp}}(x)$$

where

$$|R_i^{\text{exp}}|_{\infty} \leq K h_i^2 |\partial_y f|_{\infty} \leq C^{\text{exp}} h_i^2.$$

The claim follows by inserting (47) and the definition of $\hat{\Lambda}_{i, i+1}^x$ in (46). \square

3.1.2. Global Expansion.

Lemma 3.6.

$$\left| \Delta \hat{u}_i(x) - \Delta \hat{u}_{n-1}(x) + \sum_{j=i}^{n-2} h_j^2 Q_{t_i, t_j} \left[\hat{\Lambda}_{i, j}^x e^{h_j \eta^y(t_j, \cdot)} \varphi(t_j, \cdot) \right] (x) \right| \leq K n^{-2},$$

where $\hat{\Lambda}^x$ is defined in (45).

Proof. Let us start by finding some controls on the elements of the sum in the statement.

By independence of the increments, assuming $i + 2 \leq j \leq n - 2$ we compute

$$\mathbb{E} \left[\prod_{k=i}^{j-1} \left(1 + \eta^z(t_k, \hat{X}_{t_k}^{t_i, x}) \Delta \hat{\omega}_k \right)^2 \right] \leq \mathbb{E} \left[\prod_{k=i}^{j-2} \left(1 + \eta^z(t_k, \hat{X}_{t_k}^{t_i, x}) \Delta \hat{\omega}_k \right)^2 \right] (1 + |\eta^z(t_{j-1}, \cdot)|_{\infty} h_{j-1}),$$

so that a straightforward induction implies, for $i \leq j \leq n - 2$

$$\mathbb{E} \left[\prod_{k=i}^{j-1} \left(1 + \eta^z(t_k, \hat{X}_{t_k}^{t_i, x}) \Delta \hat{\omega}_k \right)^2 \right] \leq \exp \left(\sum_{k=i}^{j-1} |\eta^z(t_k, \cdot)|_{\infty}^2 h_k \right).$$

Hence, by the Cauchy-Schwarz theorem

$$\mathbb{E}^{\hat{\mathbb{Q}}} \left[|\hat{\Lambda}_{i, j}^x| \right] \leq \exp \left(|\eta^y|_{\infty} T + \frac{1}{2} |\eta^z|_{\infty}^2 T \right) < \infty.$$

Now, the flow property of the cubature approximation, i.e.

$$\mathbb{E}^{\mathbb{Q}} \left[\psi \left(\hat{X}_{t_k}^{t_j, \hat{X}_{t_j}^{t_i, x}} \right) \right] = \mathbb{E}^{\mathbb{Q}} \left[\psi(\hat{X}_{t_k}^{t_i, x}) \right]$$

for any measurable function ψ , implies that we get, by iterating over i the result in Lemma 3.5, that

$$(48) \quad \Delta \hat{u}_i(x) - \Delta \hat{u}_{n-1}(x) = -\mathbb{E}^{\mathbb{Q}} \left[\sum_{j=i}^{n-2} \hat{\Lambda}_{i, j}^x \left(e^{h_j \eta^y(t_j, \hat{X}_{t_j}^{t_i, x})} \varphi(t_j, \hat{X}_{t_j}^{t_i, x}) h_j^2 \right) \right] \\ + \mathbb{E}^{\mathbb{Q}} \left[\sum_{j=i}^{n-2} \hat{\Lambda}_{i, j}^x e^{h_j \eta^y(t_j, \hat{X}_{t_j}^{t_i, x})} \left(R_j^z(\hat{X}_{t_j}^{t_i, x}) - \xi(t_j, \hat{X}_{t_j}^{t_i, x}) h_j^3 \right) \right] \\ (49) \quad + \mathbb{E}^{\mathbb{Q}} \left[\sum_{j=i}^{n-2} \hat{\Lambda}_{i, j}^x e^{h_j \eta^y(t_j, \hat{X}_{t_j}^{t_i, x})} \left(h_j R_j^{u, \tilde{v}}(\hat{X}_{t_j}^{t_i, x}) - R_j^{\text{exp}}(\hat{X}_{t_j}^{t_i, x}) \Delta \hat{u}_i \right) \right].$$

Hence,

$$\begin{aligned}
& \left| \Delta \hat{u}_i(x) - \Delta \hat{u}_{n-1}(x) + \sum_{j=i}^{n-2} h_j^2 Q_{t_i, t_j} \left[\hat{\Lambda}_{i,j}^x e^{h_j \eta^y(t_j, \cdot)} \varphi(t_j, \cdot) \right] (x) \right| \\
& \leq \mathbb{E}^{\mathbb{Q}} \left[\sum_{j=i}^{n-2} \hat{\Lambda}_{i,j}^x e^{h_j \eta^y(t_j, \hat{X}_{t_j}^{t_i, x})} (||u||_{j;6,\infty} h_j^3 + C^{\exp} h_j^2 |\Delta \hat{u}_j|) \right] \\
& \quad + \mathbb{E}^{\mathbb{Q}} \left[\sum_{j=i}^{n-2} \hat{\Lambda}_{i,j}^x e^{h_j \eta^y(t_j, \hat{X}_{t_j}^{t_i, x})} (h_j (C^u |\Delta \hat{u}_j|_{\infty}^2 + C^{\tilde{v}} |\Delta \tilde{v}_j|_{\infty}^2)) \right] \\
& \leq \sum_{j=i}^{n-2} C' ||u||_{j;6,\infty} h_j^3 + \sum_{j=i}^{n-2} C_2^{\exp} h_j^2 n^{-1} \\
& \quad + \sum_{j=i}^{n-2} h_j^2 C_2^u n^{-2} + \sum_{j=i}^{n-2} h_j C_2^{\tilde{v}} |\Delta \tilde{v}_j|_{\infty}^2 \\
& \leq K n^{-2}
\end{aligned}$$

where we have used, the controls in Lemma 3.5 and the rate of convergence results in Lemma 3.4 for the first inequality, controls on the expectation of $\hat{\Lambda}_{i,j}^x e^{h_j \eta^y(t_j, \hat{X}_{t_j}^{t_i, x})}$ on the second inequality and Lemma 3.4 and Corollary A.3 for the last inequality. \square

Although in Lemma 3.6 we have identified (up to the last step) an explicit coefficient for the main error term, it still depends on the actual approximation algorithm. In order to have a more tractable expression, let us re-express the result of Lemma 3.6 in terms of a family of linear operators forming a semigroup. Let us introduce

$$(50) \quad \hat{\Theta}_{t_i, t_j} \psi(x) := Q_{t_i, t_j} \left[\hat{\Lambda}_{i,j}^x \psi(\cdot) \right] (x); \quad t_i \leq t_j.$$

where $\hat{\Lambda}_{i,j}^x$ is defined in (45).

Indeed, for any $i, j, k \in \{0, \dots, n\}$ with $i \leq j \leq k$ we can verify the associativity property

$$\hat{\Theta}_{t_i, t_j} \psi(x) \left[\hat{\Theta}_{t_j, t_k} \psi(\cdot) \right] (x) = Q_{t_i, t_j} \left[\hat{\Lambda}_{i,j}^x Q_{t_j, t_k} \left[\hat{\Lambda}_{j,k}^x \psi(\cdot) \right] \right] (x) = Q_{t_i, t_k} \left[\hat{\Lambda}_{i,k}^x \psi(\cdot) \right] (x) = \hat{\Theta}_{t_i, t_k} \psi(x)$$

and its linearity.

We show in the following that this operator can be seen as an approximation of

$$(51) \quad \Theta_{t_i, t_j} \psi(x) := P_{t_i, t_j} \left[\Lambda_{t_i, t_j}^x \psi(\cdot) \right] (x).$$

where, for $r \leq t$,

$$\begin{aligned}
(52) \quad \Lambda_{r,t}^x &= \exp \left[\int_r^t \eta^z(s, X_s^{r,x}) dW_s + \int_r^t \eta^y(s, X_s^{r,x}) - \frac{1}{2} |\eta^z(s, X_s^{r,x})|^2 ds \right], \\
&= \exp \left[\int_r^t \eta^z(s, X_s^{r,x}) \circ dW_s - \frac{1}{2} \int_r^t \left(2\eta^y(s, X_s^{r,x}) + |\eta^z(s, X_s^{r,x})|^2 + \sum_{a=1}^n V^{(a)} \eta_a^z(s, X_s^{r,x}) \right) ds \right],
\end{aligned}$$

where η^y and η^z are defined in (45). Clearly, the family $(\psi \mapsto \Lambda_{r,t}^x \psi)_{r,t}$ also forms a linear semigroup of operators.

Lemma 3.7. *Assume that $m \geq 3$, that the cubature measure is symmetric and that $\psi \in \bar{C}_b^5$. Then,*

$$|\Theta_{t_k, t_{k+1}} \psi(x) - \hat{\Theta}_{t_k, t_{k+1}} \psi(x)| \leq h_k^2 \|\psi\|_{4, \infty} + h_k^{5/2} \|\psi\|_{5, \infty}.$$

Proof. Consider the extended system

$$d \begin{bmatrix} X_t^{t_k, x} \\ \Lambda_t^{t_k, x} \\ \chi_t^{t_k, x} \end{bmatrix} = \begin{bmatrix} \bar{b}(t, X_t^{t_k, x}) \\ -\frac{1}{2} \Lambda_t^{t_k, x} \left(|\eta^z(t, X_t^{t_k, x})|^2 + \sum_{a=1}^n V^{(a)} \eta_a^z(t, X_t^{t_k, x}) \right) \\ \chi_t^{t_k, x} \left(\eta^y(t, X_t^{t_k, x}) - \eta^y(t_k, x) \right) \end{bmatrix} dt + \begin{bmatrix} \sigma(t, X_t^{t_k, x}) \\ \Lambda_t^{t_k, x} \eta^z(t, X_t^{t_k, x}) \\ 0 \end{bmatrix} \odot dW_t$$

with initial condition $(X_{t_k}^{t_k, x}, \Lambda_{t_k}^{t_k, x}, \chi_{t_k}^{t_k, x})^\top = (x, 1, e^{h_k \eta^y(t_k, x)})^\top$. Let us denote by \tilde{V}^β the associated iterated differential operators, so that we have in particular,

$$\tilde{V}^{(a)} \phi(t, y, \lambda, \chi) = V^{(a)} \phi(t, y, \lambda, \chi) + \lambda \eta_a^z(t, y) \partial_\lambda \phi(t, y, \lambda, \chi), \text{ for } a \in \{1, \dots, n\};$$

$$\begin{aligned} \tilde{V}^{(0)} \phi(t, y, \lambda, \chi) &= V^{(0)} \phi(t, y, \lambda, \chi) - \frac{1}{2} \lambda \left(|\eta^z(t, y)|^2 + \sum_{a=1}^n V^{(a)} \eta_a^z(t, y) \right) \partial_\lambda \phi(t, y, \lambda, \chi) \\ &\quad + \chi (\eta^y(t, y) - \eta^y(t_k, x)) \partial_\chi \phi(t, y, \lambda, \chi); \end{aligned}$$

$$\begin{aligned} \tilde{V}^{(a_1, a_2)} \phi(t, y, \lambda, \chi) &= V^{(a_1, a_2)} \phi(t, y, \lambda, \chi) + \lambda V^{(a_1)} [\eta_{a_2}^z](t, y) \partial_\lambda \phi(t, y, \lambda, \chi) \\ &\quad + \lambda \eta_{a_2}^z(t, y) V^{(a_1)} [\partial_\lambda \phi](t, y, \lambda, \chi) + \lambda \eta_{a_1}^z(t, y) V^{(a_2)} [\partial_\lambda \phi](t, y, \lambda, \chi) \\ &\quad + \lambda \eta_{a_1}^z(t, y) \eta_{a_2}^z(t, y) \partial_\lambda \phi(t, y, \lambda, \chi) + \lambda^2 \eta_{a_1}^z(t, y) \eta_{a_2}^z(t, y) \partial_\lambda^2 \phi(t, y, \lambda, \chi), \end{aligned}$$

for $a_1, a_2 \in \{1, \dots, n\}$.

Thus, from a Taylor-Stratonovich expansion and the expectation of iterated Stratonovich integrals, it follows that if $\Psi(t, x, \lambda, \chi) = \lambda \chi \nu(t, x)$

$$\begin{aligned} &\mathbb{E}[\Lambda_{t_{k+1}}^{t_k, x} \chi_{t_{k+1}}^{t_k, x} \nu(t_{k+1}, X_{t_{k+1}}^{t_k, x})] \\ &= e^{\eta^y(t_k, x) h_k} \nu(t_k, x) + h_k e^{\eta^y(t_k, x) h_k} \left[(V^{(0)} + \sum_{a=1}^d V^{(a, a)}) \nu \right] (t_k, x) \\ &\quad + h_k \sum_{a=1}^d e^{\eta^y(t_k, x) h_k} \eta_a^z(t_k, x) V^a [\nu](t_k, x) + \sum_{\beta \in \partial \tilde{\mathcal{A}}_3} \mathbb{E} \left(J_{t_k, t_{k+1}}^\beta \left[\tilde{V}^\beta \Psi \right] \right). \end{aligned}$$

Similarly, a development applied on ν and the properties of the symmetric cubature measure give us that

$$\begin{aligned} &\mathbb{E}^{\hat{\mathbb{Q}}}[(1 + \eta_a^z(t, x) \Delta \hat{\omega}_k) e^{\eta^y(t_k, x) h_k} \nu(t_k, \hat{X}_{t_{k+1}}^{t_k, x})] \\ &= e^{\eta^y(t_k, x) h_k} \mathbb{E}^{\hat{\mathbb{Q}}}[\nu(t_k, \hat{X}_{t_{k+1}}^{t_k, x})] + \eta_a^z(t, x) e^{\eta^y(t_k, x) h_k} \mathbb{E}^{\hat{\mathbb{Q}}}[\Delta \hat{\omega}_k \nu(t_k, \hat{X}_{t_{k+1}}^{t_k, x})] \\ &= e^{\eta^y(t_k, x) h_k} \nu(t_k, x) + h_k e^{\eta^y(t_k, x) h_k} \left[(V^{(0)} + \sum_{a=1}^d V^{(a, a)}) \nu \right] (t_k, x) \\ &\quad + h_k \eta_a^z(t_k, x) \sum_{a=1}^d e^{\eta^y(t_k, x) h_k} V^a [\nu](t_k, x) + \sum_{\beta \in \partial \tilde{\mathcal{A}}_3} \mathbb{E}^{\hat{\mathbb{Q}}} \left(I_{t_k, t_{k+1}}^\beta \left[V^\beta \nu \right] \right) \\ &\quad + \eta_a^z(t_k, x) \sum_{\beta \in \partial \tilde{\mathcal{A}}_3} \mathbb{E}^{\hat{\mathbb{Q}}} \left(I_{t_k, t_{k+1}}^{a \star \beta} \left[V^{a \star \beta} \nu \right] + \sum_{\iota=1}^{|\beta|_0} I_{t_k, t_{k+1}}^{\beta_{1:\iota} \star a \star \beta_{\iota+1:|\beta|_0}} \left[V^\beta \nu \right] \right). \end{aligned}$$

The claim follows by comparing both expansions. \square

Proposition 3.8. *Suppose that $M \geq 9$. For every $i < j \in \{1, \dots, n\}$, assume that $\gamma \geq m \geq 3$ and that the cubature measure is symmetric. Then, for φ defined in (40),*

$$\left| \sum_{j=i}^{n-2} h_j^2 \left\{ \Theta_{t_i, t_j} [\varphi(t_j, \cdot)](x) - \hat{\Theta}_{t_i, t_j} [\varphi(t_j, \cdot)](x) \right\} \right| \leq K n^{-2}.$$

Proof.

Let us write for convenience

$$\nu(t_k, x; t_j) = \Theta_{t_k, t_j} [\varphi(t_j, \cdot)](x).$$

We first show that,

$$(53) \quad |V^\beta \nu(t_k, x; t_j)| \leq K \|\varphi(t_k, \cdot)\|_{\beta, \infty} \leq K' \|u(t_k, \cdot)\|_{\beta+4, \infty}.$$

Recall that $\varphi(t_k, \cdot)$ is, essentially, a sum of derivatives of $u(t_k, \cdot)$ multiplying bounded functions. Hence, broadly speaking, the first inequality tells us that the operator regularizes as if we could “extract derivatives from the operator”. The second inequality is a direct consequence of Proposition 3.1.

It is then sufficient to prove the first inequality in (53). Since

$$\begin{aligned} \nu(t_k, x; t_j) &= \mathbb{E} \left[e^{\int_{t_k}^{t_j} \eta^z(s, X_s^{t_k, x}) dW_s + \int_{t_k}^{t_j} \{ \eta^y(s, X_s^{t_k, x}) - \frac{1}{2} |\eta^z(s, X_s^{t_k, x})|^2 \} ds} \varphi(t_j, X_{t_j}^{t_k, x}) \right] \\ &= \mathbb{E}^{\mathbb{P}^*} \left[e^{\int_{t_k}^{t_j} \eta^y(s, X_s^{t_k, x}) ds} \varphi(t_j, X_{t_j}^{t_k, x}) \right] \end{aligned}$$

for an appropriately defined \mathbb{P}^* . Given our assumptions, Girsanov’s theorem implies that the diffusion $X^{t_k, x}$ solves the SDE

$$X_t^{t_k, x} = x + \int_{t_k}^t \left\{ b(s, X_s^{t_k, x}) + \sigma(s, X_s^{t_k, x}) \eta^z(s, X_s^{t_k, x})^\top \right\} ds + \int_{t_k}^t \sigma(s, X_s^{t_k, x}) dW_s^*,$$

for a \mathbb{P}^* Brownian motion W^* , which also satisfies the structural conditions in Assumption 1.1 (resp. 1.2).

Now, note that because $\eta^y \in \mathcal{C}_b^{N-2}$, we can show that $e^{\int_{t_k}^{t_j} \eta^y(s, X_s^{t_k, x}) ds}$ is a Kusuoka-Stroock function as defined in Definition 22 in [18]. This means that the Malliavin integration by parts results hold up to multiplying by another Kusuoka-Stroock functions (Section 2.6 in [18]). We can then adapt the arguments in Corollary 32 in [18] to deduce the first inequality in (53).

Using a telescopic sum Lemma 3.7 and using (53), we conclude

$$\begin{aligned}
& \left| \sum_{j=i}^{n-2} h_j^2 \left\{ \Theta_{t_i, t_j} [\varphi(t_j, \cdot)](x) - \hat{\Theta}_{t_i, t_j} [\varphi(t_j, \cdot)](x) \right\} \right| \\
&= \left| \sum_{j=i}^{n-2} h_j^2 \sum_{k=i}^{j-1} \hat{\Theta}_{t_i, t_k} \left\{ [\hat{\Theta}_{t_k, t_{k+1}} - \Theta_{t_k, t_{k+1}}] \nu(t_{k+1}, \cdot; t_j) \right\}(x) \right| \\
&\leq \sum_{j=i}^{n-2} h_j^2 \sum_{k=i}^{j-1} \left| [\hat{\Theta}_{t_k, t_{k+1}} - \Theta_{t_k, t_{k+1}}] \nu(t_{k+1}, \cdot; t_j) \right|_{L^\infty} \\
&\leq K \sum_{j=i}^{n-2} h_j^2 \left(\sum_{k=i}^{j-1} h_k^2 \|\nu(t_{k+1}, \cdot; t_j)\|_{4, \infty} + h_k^{5/2} \|\nu(t_{k+1}, \cdot; t_j)\|_{5, \infty} \right) \\
&\leq K \sum_{j=i}^{n-2} h_j^2 \left(\sum_{k=i}^{j-1} h_k^2 \|u(t_{k+1}, \cdot)\|_{8, \infty} + h_k^{5/2} \|u(t_{k+1}, \cdot)\|_{9, \infty} \right) \\
&\leq K \frac{\gamma}{n} \sum_{j=i}^{n-2} h_j^2 \left((T - t_j)^{-3/2(1-1/\gamma)} \right) \\
&\leq K n^{-2},
\end{aligned}$$

where we used Lemma A.2 to get the last inequality. \square

Before proceeding with the proof of the main expansion result, let us give the shape of the leading coefficient appearing in the statement of the theorem.

Definition 3.9. *i. If Assumptions 1.1 holds and a uniform discretization is used,*

$$\Psi_T^{nl}(0, x) = T \int_0^T \mathbb{E}[\Lambda_{0,t}^x \varphi(t, X_t)] dt;$$

ii. If Assumptions 1.1 holds and a decreasing step discretization with $\gamma > m + 1$ is used,

$$\Psi_T^{nl}(0, x) = T^{1/\gamma} \gamma \int_0^T \mathbb{E}[\Lambda_{0,t}^x \varphi(t, X_t) (T - t)^{\frac{\gamma-1}{\gamma}}] dt;$$

where $\Lambda_{0,t}^x$ is defined in (52).

Proof of Theorem 1.4.

From Lemma 3.6, Proposition 3.8 and equation (51) we get

$$\left| \Delta \hat{u}_i(x) + \sum_{j=i}^{n-2} h_j^2 P_{t_i, t_j} \left[\Lambda_{t_i, t_j}^x \varphi(t_j, \cdot) \right](x) \right| \leq K n^{-2} + |\Delta \hat{u}_{n-1}|_\infty \leq K' n^{-2},$$

where we have used the control on the last step highlighted in Remark 3.3.

To conclude our claim, we invoke Lemma A.4 with $\psi(t) = \mathbb{E}[\Lambda_{0,t}^x \varphi(t, X_t)]$. Indeed, the control on $|\psi|$ readily follows from (42) and the expression for $\Lambda_{0,t}$. Similarly, using the chain rule and the equality

$$\partial_t \mathbb{E}[V^\beta u(t, X_t)] = \mathbb{E} \left[V^{0\star\beta} u(t, X_t) + \sum_{i=1}^d V^{(i,i)\star\beta} u(t, X_t) \right],$$

we conclude that ψ has well defined locally bounded first order derivatives in $[0, T)$, and hence it is of bounded variation in $[0, T - \epsilon]$ for all $\epsilon > 0$. \square

4. COMPLEXITY REDUCTION

In this section, we aim to control the complexity growth on both the number of steps n and the dimension of the problem. To achieve this, we make use of the regularity of the field u to find a sparse representation of the solution.

Importantly, we work here under Assumption 1.5 which is stronger than the UFG condition used in the previous sections. It implies that all partial derivatives of the solution of the backward equation are well defined. Moreover, it implies that we are able to control any partial derivative of the solution of the backward equation, say $\partial_i u$, using the control on the derivative operators $V^i u$ modulo a constant. Hence, we can use Propositions 2.6 and 3.1 to control any partial derivative of the function u .

The strong regularity of the backward solution can be used to obtain accurate representations of the function using the information in a reduced number of points. We examine here the sparse grid technique (cf. [7]) to achieve this goal.

Sparse grid are obtained from expressing a function in terms of a truncated hierarchical basis satisfying the condition that the truncation error and the growth of the hierarchical basis support can be controlled to obtain an efficient representation of the function. In the particular case of hierarchical basis based on linear interpolation, the control on the approximation is shown assuming a bound on all mixed derivatives of a given order, which is an available result in our setup.

In this section, we study the combination of the sparse grid technique based on linear interpolation with the previously presented scheme to solve BSDEs. This will allow us to profit more effectively of a possible extrapolation, as we will show in the following.

4.1. Sparse representation operator. We discuss now some key properties of the sparse representation that was introduced in Section 1.4.2.

First, It can be easily verified that, in the case $d = 1$, $\pi_{\mathcal{V}_p}^A[\psi](\cdot)$, recall (12)-(13), is the usual linear interpolation of ψ along a uniform grid with 2^{p+1} points (see Lemma 4.2 below). However, as the number of dimensions increases, the operator $\pi_{\mathcal{V}_p}^A[\cdot]$ diverges from the usual linear interpolation: not all points in the uniform grid are included. Moreover, under certain assumptions on the function to be approximated, the induced maximum error is just a log factor worst than the one obtained via a uniform grid. This is proved in [7] (Lemma 3.6 and Theorem 3.8) and is presented here for the readers convenience.

Proposition 4.1. *Let ψ defined in A such that $D^\beta \psi$ is bounded Lebesgue-almost surely in A for all multi-index $|\alpha|_\infty < 2$. Let $p \geq d - 1$. Then, there exists constants K and K' such that for all $x \in A$,*

$$|\psi - \pi_p^A \psi(x)| \leq 2^{-2p} K 2^{-d+1} \frac{(p - d + 1)^{d-1}}{(d - 1)!} |D^{\mathbf{2}} \psi|_\infty$$

and

$$N_{p,d}^{Total} := \# \{ \tilde{\mathbf{x}}_{\mathbf{l},j} : (\mathbf{l}, j) \in \mathcal{I}_p(A) \} \leq 2^p K' 2^{-d+1} \frac{(p - d + 1)^{d-1}}{(d - 1)!}.$$

We also deduce the following properties of the operator $\pi_{\mathcal{V}_p}^A$.

Lemma 4.2. *For $a < b \in \mathbb{R}$, let $\phi : [a, b] \rightarrow \mathbb{R}$. Let*

$$\bar{\pi}_m(\psi; [a, b])(x) := \sum_{0 \leq j \leq 2^m} \psi(\tilde{x}_{m,j}) \phi_{m,j}(x)$$

i.e., $\bar{\pi}_m$ is defined as the usual linear interpolation operator in dimension one, with 2^m points.

Then, for $\psi : A = \prod_{i=1}^d [a_i, b_i] \rightarrow \mathbb{R}$, we have the representation, for $d > 1$,

$$\pi_{\mathcal{V}_p}^A[\psi](x) = \sum_{k=0}^p \bar{\pi}_k \left(y \mapsto \pi_{\mathcal{V}_{p-k}}^{A-}[\psi(\cdot, y)](x_1, \dots, x_{d-1}); [a_d, b_d] \right) (x_d)$$

and, for $d = 1$,

$$\pi_{\mathcal{V}_p}^{[a,b]}[\psi](x) = \bar{\pi}_p(\psi; [a, b])(x) .$$

Proof. We consider first the case $d = 1$. In this case we prove that $\pi_{\mathcal{V}_m}^{[a,b]}[\psi](x) = \bar{\pi}_m(\psi, [a, b])(x)$, by induction on m . Note that if $m = 0$, the claim is trivial. For $m > 0$, we note that linearity implies, for a fixed $m > 0$ and $0 < k \leq 2^{m-1}$

$$(54) \quad \phi_{m-1,k}(x) = \phi_{m,2k}(x) + \frac{1}{2}\phi_{m,2k-1}(x) + \frac{1}{2}\phi_{m,2k+1}(x).$$

Since $\check{x}_{m,2k} = \check{x}_{m-1,k}$, we deduce from (13), (54) and using induction,

$$(55) \quad \begin{aligned} \pi_{\mathcal{V}_m}^{[a,b]}[\psi](x) &= \sum_{l,j \in \mathcal{I}_m(A)} \theta_{l,j}(\psi, [a, b]) \phi_{l,j}(x) \\ &= \pi_{\mathcal{V}_{m-1}}^{[a,b]}[\psi](x) + \sum_{0 \leq k \leq 2^{m-1}} \left[\psi(\check{x}_{m,2k+1}) - \frac{1}{2}\psi(\check{x}_{m,2k}) - \frac{1}{2}\psi(\check{x}_{m,2k+2}) \right] \phi_{m,2k+1}(x) \\ &= \bar{\pi}_{m-1}(\psi; [a, b])(x) + \sum_{0 \leq k \leq 2^{m-1}} \left[\psi(\check{x}_{m,2k+1}) - \frac{1}{2}\psi(\check{x}_{m,2k}) - \frac{1}{2}\psi(\check{x}_{m,2k+2}) \right] \phi_{m,2k+1}(x) \\ &= \sum_{0 \leq k \leq 2^{m-1}} \psi(\check{x}_{m-1,k}) \phi_{m-1,k}(x) \\ &\quad + \sum_{0 \leq k \leq 2^{m-1}} \left[\psi(\check{x}_{m,2k+1}) - \frac{1}{2}\psi(\check{x}_{m,2k}) - \frac{1}{2}\psi(\check{x}_{m,2k+2}) \right] \phi_{m,2k+1}(x) \\ &= \sum_{0 \leq k \leq 2^{m-1}} \psi(\check{x}_{m,2k}) \left[\phi_{m,2k}(x) + \frac{1}{2}\phi_{m,2k-1}(x) + \frac{1}{2}\phi_{m,2k+1}(x) \right] \\ &\quad + \sum_{0 \leq k \leq 2^{m-1}} \left[\psi(\check{x}_{m,2k+1}) - \frac{1}{2}\psi(\check{x}_{m,2k}) - \frac{1}{2}\psi(\check{x}_{m,2k+2}) \right] \phi_{m,2k+1}(x) \\ &= \bar{\pi}_m(\psi; [a, b])(x). \end{aligned}$$

We now consider the case $d > 1$. From the sparse operator definition in (12), (11) and (10), we observe

$$\pi_{\mathcal{V}_p}^A[\psi](x) = \sum_{k=0}^p \sum_{(l_d, j_d) \in \mathcal{I}_k} \sum_{(\tilde{l}, \tilde{j}) \in \mathcal{I}_{p-k}(A-)} \theta_{\tilde{l}*(l_d), \tilde{j}*(j_d)}(\psi, A) \phi_{\tilde{l}, \tilde{j}}((x_1, \dots, x_{d-1}); A-) \phi_{l_d, j_d}(x_d; [a_d, b_d]) .$$

Using the definition (13) and reordering, the above reads

$$\begin{aligned} \pi_{\mathcal{V}_p}^A[\psi](x) &= \sum_{k=0}^p \sum_{(l_d, j_d) \in \mathcal{I}_k} \phi_{l_d, j_d}(x_d, [a_d, b_d]) \left(\pi_{\mathcal{V}_{p-k}}^{A-}[\psi(\cdot, x_{l_d, j_d}^d)](x) 1, \dots, x) d-1 \right) \\ &\quad - \frac{1_{\{l_d \neq 0\}}}{2} \pi_{\mathcal{V}_{p-l_d}}^{A-}[\psi(\cdot, x_{l_d, j_d-1}^d)](x_1, \dots, x_{d-1}) - \frac{1_{\{l_d \neq 0\}}}{2} \pi_{\mathcal{V}_{p-l_d}}^{A-}[\psi(\cdot, x_{l_d, j_d+1}^d)](x_1, \dots, x_{d-1}) \Bigg) . \end{aligned}$$

For the last sum, we recognise the sparse approximation at level k of $y \mapsto \pi_{\mathcal{V}_{p-d}}^{A-}[\psi(\cdot, y)](x_1, \dots, x_{d-1})$ evaluated at x_d . The proof is then concluded using the characterisation obtained in the one-dimensional case. \square

The previous representation implies some important stability properties of the sparse interpolation operator.

Corollary 4.3. *For any $p > 0$ and all $x \in A$, we have*

- i. $\pi_{\mathcal{V}_p}^A[1](x) = 1$;
- ii. For all $\alpha_1, \alpha_2 \in \mathbb{R}$, $\pi_{\mathcal{V}_p}^A[[\alpha_1\psi_1 + \alpha_2\psi_2]](x) = \alpha_1\pi_{\mathcal{V}_p}^A[\psi_1](x) + \alpha_2\pi_{\mathcal{V}_p}^A[\psi_2](x)$;
- iii. If $\psi \geq 0$, $\pi_{\mathcal{V}_p}^A\psi \geq 0$;
- iv. $\pi_{\mathcal{V}_p}^A[\psi](x) \leq |\psi|_\infty$.

Proof. Claims i., ii. and iii. follow from an induction procedure using Lemma 4.2 and the analogous properties of the usual interpolation operator.

To prove iv. we have the following chain of implications for all $x \in A$:

$$\begin{aligned} |\psi|_\infty - \psi(x) \geq 0 &\Rightarrow \pi_{\mathcal{V}_p}^A[(|\psi|_\infty - \psi)](x) \geq 0 \\ &\Rightarrow \pi_{\mathcal{V}_p}^A[|\psi|_\infty](x) - \pi_{\mathcal{V}_p}^A[\psi](x) \geq 0 \\ &\Rightarrow |\psi|_\infty \geq \pi_{\mathcal{V}_p}^A[\psi](x). \end{aligned}$$

\square

In the next section, we turn to the study of the scheme presented in (16)-(17). The effect of adding the projection operator means that we do not need to construct the cubature tree as before, but only to construct the increments starting from nodes in the sets $(D_i)_{i=0, \dots, n-1}$, recall (14). Note that we are not projecting at the final time $t_n = T$. This is in order to be able to profit from the regularization property when the condition is only Lipschitz. Based on the results in [7] we show in Theorem 4.5 that this reduces the complexity requirements of the algorithm.

In practice, one can decide to only use the projection on sparse grids whenever the number of nodes in the cubature tree is bigger than a given threshold. However, to ease the analysis, we assume the projection is done at each step of the algorithm, and we ignore the potential advantage of adding this conditioning on the projection.

4.2. Error analysis and proof of complexity. The additional projection term induces some adjustments in the error analysis of Section 3. The error we study now is defined by

$$\check{\mathcal{E}}_i(x) := \check{u}_i(x) - u(t_i, x).$$

This time, the role of \tilde{v} is played by

$$\ddot{v}_i(x) := \pi_{\mathcal{V}_{p-d+1}}^{D_i} \left[\left(\mathbb{E}^{\hat{\mathbb{Q}}} \left[u(t_{i+1}, \hat{X}_{t_{i+1}}^{t_i, \cdot}) \frac{\Delta \hat{w}_i}{h_i} \right] \right) \right](x),$$

and as before, we define

$$u(t_i, x) = \pi_{\mathcal{V}_{p-d+1}}^{D_i} \left[\left(\mathbb{E}^{\hat{\mathbb{Q}}} \left[u(t_{i+1}, \hat{X}_{t_{i+1}}^{t_i, \cdot}) \right] \right) \right](x) + h_i f(x, u(t_i, x), \ddot{v}_i(x)) + \check{\xi}(x).$$

Regarding the one step-controls given in Lemma 3.2, controls i. and ii. are adjusted as follows:

Lemma 4.4 (One step controls). *Let $\hat{\mathbb{Q}}$ be a cubature measure constructed from a symmetric cubature formula of order $m \geq 3$. Under Assumptions 1.1 or 1.2 with $M > (2d) \vee (m+3)$ and Assumption 1.5, we have, for all $x \in D_i$,*

$$\begin{aligned} & \left| \mathbb{E} \left[u(t_{i+1}, X_{t_{i+1}}^{t_i, x}) \right] - \pi_{\mathcal{V}_{p_i-d+1}}^{D_i} \left[\mathbb{E}^{\mathbb{Q}} \left[u(t_{i+1}, \hat{X}_{t_{i+1}}^{t_i, \cdot}) \right] \right] (x) \right| \\ & \leq K' (2^{-2p} (p-d+1)^{d-1} \|u(t_{i+1}, \cdot)\|_{2d, \infty} + h_i^{\frac{m+1}{2}} \|u\|_{i; m+1, \infty} + h_i^{\frac{m+3}{2}} \|u\|_{i; m+3, \infty}) \end{aligned}$$

and, if $M > (2d+1) \vee (m+1)$

$$\begin{aligned} & |f(t_i, x, u(t_i, x), \check{v}_i(x)) - f(t_i, x, u(t_i, x), \check{v}_i(x))| \\ & \leq K' (2^{-2p} (p-d+1)^{d-1} \|u(t_{i+1}, \cdot)\|_{2d+1, \infty} + h_i^{\frac{m-1}{2}} \|u\|_{i; m, \infty} + h_i^{\frac{m+1}{2}} \|u\|_{i; m+2, \infty}). \end{aligned}$$

Proof. We have

$$\begin{aligned} & \left| \mathbb{E} \left[u(t_{i+1}, X_{t_{i+1}}^{t_i, x}) \right] - \pi_{\mathcal{V}_{p_i-d+1}}^{D_i} \left[\mathbb{E}^{\mathbb{Q}} \left[u(t_{i+1}, \hat{X}_{t_{i+1}}^{t_i, \cdot}) \right] \right] (x) \right| \\ & \leq \left| \mathbb{E} \left[u(t_{i+1}, X_{t_{i+1}}^{t_i, x}) \right] - \pi_{\mathcal{V}_{p_i-d+1}}^{D_i} \left[\mathbb{E} \left[u(t_{i+1}, X_{t_{i+1}}^{t_i, \cdot}) \right] \right] (x) \right| \\ & \quad + \left| \pi_{\mathcal{V}_{p_i-d+1}}^{D_i} \left[\left(\mathbb{E} \left[u(t_{i+1}, X_{t_{i+1}}^{t_i, x}) \right] - \mathbb{E}^{\mathbb{Q}} \left[u(t_{i+1}, \hat{X}_{t_{i+1}}^{t_i, \cdot}) \right] \right) \right] (x) \right| \\ & \leq 2^{-(2p+3d-1)} K \frac{(p-d+1)^{d-1}}{(d-1)!} \left| D^2 \mathbb{E} \left[u(t_{i+1}, X_{t_{i+1}}^{t_i, \cdot}) \right] \right|_{\infty} \\ & \quad + K m h_i^{\frac{m+1}{2}} \|u\|_{i; m+1, \infty} + h_i^{\frac{m+3}{2}} \|u\|_{i; m+3, \infty} \\ & \leq K' (2^{-2p} (p-d+1)^{d-1} \|u(t_{i+1}, \cdot)\|_{2d, \infty} + h_i^{\frac{m+1}{2}} \|u\|_{i; m+1, \infty} + h_i^{\frac{m+3}{2}} \|u\|_{i; m+3, \infty}) \end{aligned}$$

where the first term in the second inequality comes from Proposition 4.1, and the second term from Proposition 2.7 and Lemma 4.3.

By a similar development, we have

$$\begin{aligned} & |f(t_i, x, u(t_i, x), \check{v}_i(x)) - f(t_i, x, u(t_i, x), \check{v}_i(x))| \\ & \leq \left| \partial_z f \right|_{\infty} \left| \mathbb{E} \left[u(t_{i+1}, X_{t_{i+1}}^{t_i, x}) \frac{\Delta W_{t_i}}{h_i} \right] - \pi_{\mathcal{V}_{p_i-d+1}}^{D_i} \left[\mathbb{E}^{\mathbb{Q}} \left[u(t_{i+1}, \hat{X}_{t_{i+1}}^{t_i, \cdot}) \frac{\Delta \hat{W}_i}{h_i} \right] \right] (x) \right| \\ & \leq \left| \partial_z f \right|_{\infty} \left\{ \left| \mathbb{E} \left[u(t_{i+1}, X_{t_{i+1}}^{t_i, x}) \frac{\Delta W_{t_i}}{h_i} \right] - \pi_{\mathcal{V}_{p_i-d+1}}^{D_i} \left[\mathbb{E} \left[u(t_{i+1}, X_{t_{i+1}}^{t_i, \cdot}) \frac{\Delta W_{t_i}}{h_i} \right] \right] (x) \right| \right. \\ & \quad \left. \left| \pi_{\mathcal{V}_{p_i-d+1}}^{D_i} \left[\mathbb{E} \left[u(t_{i+1}, X_{t_{i+1}}^{t_i, \cdot}) \frac{\Delta W_{t_i}}{h_i} \right] \right] (x) - \pi_{\mathcal{V}_{p_i-d+1}}^{D_i} \left[\mathbb{E}^{\mathbb{Q}} \left[u(t_{i+1}, \hat{X}_{t_{i+1}}^{t_i, \cdot}) \frac{\Delta \hat{W}_i}{h_i} \right] \right] (x) \right| \right\} \\ & \leq 2^{-(2p+3d-1)} K \frac{(p-d+1)^{d-1}}{(d-1)!} \left| D^2 \mathbb{E} \left[u(t_{i+1}, X_{t_{i+1}}^{t_i, x}) \frac{\Delta W_{t_i}}{h_i} \right] \right|_{\infty} \\ & \quad + K h_i^{\frac{m-1}{2}} \|u(t_i, \cdot)\|_{m, \infty} + h_i^{\frac{m+1}{2}} \|u\|_{i; m+2, \infty} \\ & \leq K' (2^{-2p} (p-d+1)^{d-1} \|u(t_{i+1}, \cdot)\|_{2d+1, \infty} + h_i^{\frac{m-1}{2}} \|u\|_{i; m, \infty} + h_i^{\frac{m+1}{2}} \|u\|_{i; m+2, \infty}). \end{aligned}$$

□

Note that this time we only have an upper bound and not an explicit one-step expansion.

Finally, the proof of Theorem 1.4 can be reproduced by using the new controls.

Proposition 4.5. *Suppose that Assumption 1.5 and either Assumption 1.1 or 1.2 hold with $M > (2d+1) \vee 9$. Define*

$$m^* = \begin{cases} \frac{m+1}{2} & \text{under Assumption 1.1} \\ \left(d + \frac{m-1-\gamma}{2\gamma} \right) \vee \frac{m+1}{2} & \text{under Assumption 1.2} \end{cases};$$

and let (p_0, \dots, p_n) be given by

$$p_i = \inf\{a \in \mathbb{N} : 2a - (d-1) \log_2(a-d+1) > -m^* \log_2 h_i; a > d-1\}$$

Then, the claim of theorem 1.4 holds. Moreover, there exists $K > 0$ independent of d , such that

$$\#\mathcal{D} \leq 1 + K \sum_{i=1}^{n-2} h_i^{-m^*} |m^* \log_2 h_i|^{d-1} + K(1 + \kappa) h_{n-1}^{-m^*} |m^* \log_2 h_{n-1}|^{d-1},$$

recall (15).

Proof. From the assumption of the proposition,

$$2^{-2p_i} (p_i - d + 1)^{d-1} < \begin{cases} h_i^{\frac{m+1}{2}} & \text{under Assumption 1.1} \\ h_i^{m^*} & \text{under Assumption 1.2} \end{cases}$$

and hence, from Lemma 4.4

$$\begin{aligned} & \left| \mathbb{E} \left[u(t_{i+1}, X_{t_{i+1}}^{t_i, x}) \right] - \pi_{\mathcal{V}_{p_i-d+1}}^{D_i} \left[\mathbb{E}^{\mathbb{Q}} \left[u(t_i, \hat{X}_{t_{i+1}}^{t_i, \cdot}) \right] \right] (x) \right| \\ & \leq \begin{cases} h_i^{\frac{m+1}{2}} (\|u(t_{i+1}, \cdot)\|_{2d, \infty} + \|u\|_{i; m+1, \infty} + h_i \|u\|_{i; m+3, \infty}) & \text{A. 1.1} \\ (h_i^{d+(m-1-\gamma)/(2\gamma)} \wedge h_i^{\frac{m+1}{2}}) \|u(t_{i+1}, \cdot)\|_{2d, \infty} + h_i^2 \|u\|_{i; m+1, \infty} + h_i^3 \|u\|_{i; m+3, \infty} & \text{A. 1.2} \end{cases} \end{aligned}$$

Analogously

$$\begin{aligned} & |f(t_i, x, u(t_i, x), \check{v}_i(x)) - f(t_i, x, u(t_i, x), \check{v}_i(x))| \\ & \leq \begin{cases} h_i (\|u(t_{i+1}, \cdot)\|_{2d+1, \infty} + \|u\|_{i; m, \infty} + h_i \|u\|_{i; m+2, \infty}) & \text{A. 1.1} \\ (h_i^{d+(m-1-\gamma)/(2\gamma)} \wedge h_i^{\frac{m+1}{2}}) \|u(t_i, \cdot)\|_{2d+1, \infty} + h_i \|u\|_{i; m, \infty} + h_i^2 \|u\|_{i; m+2, \infty} & \text{A. 1.2} \end{cases} \end{aligned}$$

We conclude following the lines of the proof of Theorem 1.4 to get the first part of the claim, given that, by Lemma A.4, the total contribution of the first term in each case is of higher order.

To prove the result on the complexity growth, note that, by the minimality property of p_i ,

$$2(p_i - 1) - (d-1) \log_2(p_i - d) \leq -m^* \log_2(h_i).$$

But, since

$$2(p_i - 1) - (d-1) \log_2(p_i - d) = (p_i - d + 1) + [(p_i - 2) + (d-1)(1 - \log_2(p_i - d))] \geq p_i - d + 1$$

for p_i large enough, we get

$$2^{p_i} (p_i - d + 1)^{d-1} \leq h_i^{-m^*} (-m^* \log_2(h_i))^{2d-2}.$$

The result follows from Proposition 4.1, and the fact that no sparse grid is used at time $t_n = T$. \square

As a Corollary, we can conclude this section by giving the proof of Theorem 1.6 which analyses the complexity of the approximation.

Proof of Theorem 1.6

i. By the explicit error expansion in Theorem 1.4,

$$N = K' \Delta^{-1}$$

for some constant $K > 0$. The claim follows directly from Proposition 4.5 in this case, modulo recalling that for any $\epsilon > 0$ there is a sufficiently large constant to control the $|\log(h_n)|^{d-1}$ term.

- ii. By the explicit error expansion in Theorem 1.4, and given the assumed symmetry of the cubature formula,

$$N \leq K' \Delta^{-1/2}$$

for some constant $K > 0$. Then, the claim follows from Corollary A.5, which shows that the particular form of the extrapolation produces the same limit error, and Proposition 4.5. \square

Proposition 4.5 and Theorem 1.6 shows that, for a fixed dimension, the complexity of the algorithm in terms of the number of discretization steps n is at most polynomial. Moreover, the dependence with respect to the dimension is constant in the smooth boundary condition case of Assumption (1.1) and depending on n under Assumption (1.2). We also conclude that, in all the analyzed cases, the explicit expansion allows for an extrapolation procedure that effectively reduces the asymptotic convergence of the algorithm.

4.2.1. Excluding zero-level points. Let us consider the situation whereby we are interested in obtaining only an estimation of the value Y_0^x .

We can use well-known control results under an uniform ellipticity assumption to eliminate nodes with relatively low probability (more specifically nodes belonging to multi-levels containing zeros). We work then assuming that there exists $\Lambda > 0$ such that

$$\frac{1}{\Lambda} |x|^2 \leq x^\dagger \sigma \sigma^\dagger x \leq \Lambda |x|^2; \quad \text{for all } x \in \mathbb{R}^d.$$

Let us start by recalling that a well known Malliavin calculus result shows the existence of a family of transition densities of the process X denoted by $p_X(t_1, x_1; t_2, x_2)$ under the uniform ellipticity assumption. Moreover, this density is bounded below and above by Gaussian densities: indeed, one can prove that there exist constants $C_-, C_+, \sigma_-, \sigma_+$ such that if

$$\varphi(x; \sigma^2) = \frac{1}{(2\pi\sigma^2)^{d/2}} e^{-\frac{|x|^2}{2\sigma^2}}$$

then

$$(56) \quad C_- \varphi(|x_2 - x_1|, (t_2 - t_1)\sigma_-^2) \leq p_X(t_1, x_1; t_2, x_2) \leq C_+ \varphi(|x_2 - x_1|, (t_2 - t_1)\sigma_+^2).$$

The upper bound in equation (56) is proved in [21] using the parametrix method, while the lower bound under our regularity assumptions is shown in [1].

Using the stated Gaussian bounds and the fact that for a Gaussian random variable $L \sim \mathcal{N}(m, \sigma^2)$, $\mathbb{P}(|L - m| \geq y) \leq e^{-\frac{y^2}{2\sigma^2}}$, we can deduce that there exists a family of sets A_1, \dots, A_n , such that

$$\mathbb{P}(X_{t_k} \notin A_k) \leq h_k^3$$

and,

$$\text{diam}(A_k) \leq K t_k \log^{1/2}(h_k^{-1}).$$

Hence, for any bounded measurable function ψ ,

$$|\mathbb{E}[\psi(X_{t_k})] - \mathbb{E}[\psi(X_{t_k}) \mathbf{1}_{\{X_{t_k} \in A_k\}}]| \leq |\psi|_\infty h_k^3.$$

Since, we have that

$$Y_0^x = \mathbb{E}[g(X_T)] + \sum_{k=0}^{n-1} \mathbb{E} \left[\int_{t_k}^{t_{k+1}} f(s, X_s, Y_s, Z_s) ds \right],$$

we get that the solution of the “tamed” BSDE

$$\check{Y}_0^x = \mathbb{E}[g(X_T) \mathbf{1}_{X_T \in A_n}] + \sum_{k=0}^{n-1} \mathbb{E} \left[\mathbf{1}_{X_{t_k} \in A_k} \int_{t_k}^{t_{k+1}} f(s, X_s, Y_s, Z_s) ds \right]$$

satisfies

$$|Y_0^x - \check{Y}_0^x| \leq K h_k^2.$$

Therefore, instead of solving the original problem, we can solve the “tamed” BSDE and the error would be controlled at the initial time. On the other hand, the uniform-ellipticity assumption on the diffusion coefficients, and matrix norm equivalence implies that there exist two constants, λ_- , λ_+ such that

$$\lambda_- \sum_{k=0}^i h_k^{1/2} \leq \text{diam}[(\text{supp}(\mathbb{Q}_{t_i}))^\blacksquare] \leq \lambda_+ \sum_{k=0}^i h_k^{1/2}$$

Since for $h_k < 1$ we get $h_k^{1/2} > h_k$. We conclude that there exists \hat{n} large enough so that, for all $n > \hat{n}$, $A_n \subset (\text{supp}(\mathbb{Q}_{t_n}))^\blacksquare$. In particular, the points in the boundary of the support will have value zero for the tamed BSDE, and we can exclude them from the sparse grid representation without changing the rate of convergence of approximation of Y_0^x .

If the intermediate values of the backward variables are needed, an improvement in complexity can be obtained by using either non-homogeneous grids or coarser basis elements towards the extremes. However, we do not consider the detailed analysis of these modifications in this work.

APPENDIX A. AUXILIARY RESULTS ON THE DECREASING STEP DISCRETIZATION

Proposition A.1. *Let $a > 1$, $n > 0$. There exists $C > 0$ such that*

$$\frac{1}{n^a} \sum_{k=1}^n \left(\frac{k}{n}\right)^b \leq \begin{cases} C n^{1-a} & \text{if } b > -1 \\ C n^{1-a} \log(n) & \text{if } b = -1 \\ C n^{-(a+b)} & \text{if } b < -1 \end{cases}.$$

Proof.

- If $b \geq 0$, the function x^b is non-decreasing in $[0, 1]$ and

$$\frac{1}{n} \sum_{k=1}^n \left(\frac{k}{n}\right)^b \leq \int_{1/n}^1 x^b dx \leq \frac{1}{b+1} (1 - n^{-(b+1)}) \leq (b+1)^{-1}.$$

- If $-1 \leq b < 0$, the function x^b is decreasing in $[0, 1]$ and

$$\frac{1}{n} \sum_{k=1}^n \left(\frac{k}{n}\right)^b = n^{-(b+1)} + \frac{1}{n} \sum_{k=1}^{n-1} \left(\frac{k}{n}\right)^b \leq 1 + \int_{1/n}^{1-1/n} x^b dx,$$

where the last integral is bounded by $(b+1)^{-1}$ if $b > -1$ and by $\log(n)$ if $b = -1$.

- If $b < -1$, the series $\sum_{k=1}^n k^b$ is convergent and increasing. The claim follows with C the limit of the series.

Lemma A.2. *Let $\gamma \geq 1, \ell \geq 1$. Let ψ be such that in $[0, T)$,*

$$|\psi(t)| \leq C |T - t|^{-\beta}$$

Then, there exists a $C' > 0$ such that

$$(57) \quad \sum_{k=0}^{n-1} |\psi(t_k)| h_k^\ell \leq \begin{cases} C' n^{-(\ell-1)} & \text{if } \gamma(\ell - \beta) > \ell - 1 \\ C' n^{-(\ell-1)} \log(n) & \text{if } \gamma(\ell - \beta) = \ell - 1 \\ C' n^{-\gamma(\ell-\beta)} & \text{if } \gamma(\ell - \beta) < \ell - 1 \end{cases}.$$

Proof. Remark that

$$h_k = T\gamma \int_{\frac{k}{n}}^{\frac{k+1}{n}} (1-x)^{\gamma-1} dx$$

and for $\gamma \geq 1$, $(1-x)^{\gamma-1}$ is decreasing in $[0, 1]$. This implies

$$(58) \quad \frac{T\gamma}{n} \left(1 - \frac{k+1}{n}\right)^{\gamma-1} \leq h_k \leq \frac{T\gamma}{n} \left(1 - \frac{k}{n}\right)^{\gamma-1}.$$

Thus, we obtain

$$\begin{aligned} \sum_{k=0}^{n-1} |\psi(t_k)| h_k^\ell &\leq \left(\frac{T\gamma}{n}\right)^\ell \sum_{k=0}^{n-1} |T - t_k|^{-\beta} \left(1 - \frac{k}{n}\right)^{\ell(\gamma-1)} \\ &= \left(\frac{T\gamma}{n}\right)^\ell \sum_{k=0}^{n-1} \left(1 - \frac{k}{n}\right)^{\ell(\gamma-1)-\beta\gamma} \\ &= \left(\frac{T\gamma}{n}\right)^\ell \sum_{k=1}^n \left(\frac{k}{n}\right)^{\ell(\gamma-1)-\beta\gamma}. \end{aligned}$$

We conclude from Proposition A.1. \square

Corollary A.3. *With the assumptions of Lemma A.2,*

$$\sum_{k=0}^{n-2} \sup_{t \in [t_k, t_{k+1}]} |\psi(t)| h_k^\ell \leq \begin{cases} C' n^{-(\ell-1)} & \text{if } \gamma(\ell - \beta) > \ell - 1 \\ C' n^{-(\ell-1)} \log(n) & \text{if } \gamma(\ell - \beta) = \ell - 1 \\ C' n^{-\gamma(\ell-\beta)} & \text{if } \gamma(\ell - \beta) < \ell - 1 \end{cases}.$$

Proof. If $\beta \leq 0$, the control is decreasing and the claim follows directly from Lemma A.2.

Otherwise, we observe

$$\begin{aligned} \sum_{k=0}^{n-2} \sup_{t \in [t_k, t_{k+1}]} |\psi(t)| h_k^\ell &\leq \sum_{k=0}^{n-2} \left(1 - \frac{k+1}{n}\right)^{-\gamma\beta} h_k^\ell \\ &\leq \left(\frac{T\gamma}{n}\right)^\ell \sum_{k=0}^{n-2} \left(1 - \frac{k+1}{n}\right)^{-\gamma\beta} \left(1 - \frac{k}{n}\right)^{\ell(\gamma-1)} \\ &\leq \left(\frac{T\gamma}{n}\right)^\ell 2^{\ell(\gamma-1)} \sum_{k=1}^{n-1} \left(\frac{k}{n}\right)^{-\gamma\beta+\ell(\gamma-1)} \end{aligned}$$

and we conclude using Proposition A.1 \square

Lemma A.4. *Under the same assumptions of Lemma A.2, suppose that $\gamma(\ell - \beta) > \ell$ and that ψ is of bounded variation in $[0, T - \epsilon]$ for all $\epsilon > 0$. Then, there exists $K > 0$ such that*

$$(59) \quad \sum_{k=0}^{n-1} \psi(t_k) h_k^\ell = \left(\frac{T\gamma}{n}\right)^{\ell-1} \int_0^T \psi(t) \left(1 - \frac{t}{T}\right)^{\beta^*} dt + K n^{-\ell}.$$

where

$$(60) \quad \beta^* := (\ell - 1)(1 - \gamma^{-1}).$$

Proof. First, we observe that

$$\int_0^T \psi(t) \left(1 - \frac{t}{T}\right)^{\beta^*} dt = T\gamma \int_0^1 \psi(T\{1 - (1-x)^\gamma\})(1-x)^{\gamma(\beta^*+1)-1} dx =: \int_0^1 \theta(x) dx ,$$

and since $\gamma(\ell - \beta) > \ell$

$$(61) \quad |\theta(x)| \leq C(1-x)^{\gamma(\beta^*-\beta+1)-1} = C(1-x)^{\gamma(\ell-\beta)-\ell} .$$

We compute that

$$\sum_{k=0}^{n-1} \psi(t_k) h_k^\ell = \frac{(T\gamma)^\ell}{n^{\ell-1}} \int_0^1 \theta(x) dx + R_n^2 + R_n^1 ,$$

with

$$R_n^1 := \left(\frac{T\gamma}{n}\right)^\ell \sum_{k=0}^{n-1} \psi(t_k) \left(\left(\frac{n}{T\gamma}\right)^\ell h_k^\ell - \left(1 - \frac{k}{n}\right)^{\ell(\gamma-1)} \right) ,$$

$$R_n^2 := \frac{(T\gamma)^\ell}{n^{\ell-1}} \left(\sum_{k=0}^{n-1} \frac{\psi(t_k)}{n} \left(1 - \frac{k}{n}\right)^{\ell(\gamma-1)} - \int_0^1 \theta(x) dx \right) .$$

We now study each remainder term separately.

i. We observe that, from (58), we have

$$(62) \quad \left| \left(1 - \frac{k}{n}\right)^{\ell(\gamma-1)} - \left(\frac{n}{T\gamma}\right)^\ell h_k^\ell \right| \leq \left| \left(1 - \frac{k}{n}\right)^{\ell(\gamma-1)} - \left(1 - \frac{k+1}{n}\right)^{\ell(\gamma-1)} \right|$$

$$\leq \frac{\ell(\gamma-1)}{n} \left(1 - \frac{k}{n}\right)^{\ell(\gamma-1)-1} ,$$

since $\ell(\gamma-1) \geq 1$.

Using (61) and (62)

$$|R_n^1| \leq \left(\frac{T\gamma}{n}\right)^\ell \frac{\ell(\gamma-1)}{n} \sum_{k=0}^{n-1} \left| \theta\left(\frac{k}{n}\right) \right| \left(1 - \frac{k}{n}\right)^{-1}$$

$$\leq \left(\frac{T\gamma}{n}\right)^\ell \frac{\ell(\gamma-1)}{n} \sum_{k=1}^n \left(\frac{k}{n}\right)^{\gamma(\ell-\beta)-\ell-1}$$

$$\leq \frac{K}{n^\ell} ,$$

where we used, for the last inequality Proposition A.1, and the fact that $\gamma(\ell - \beta) > \ell$.

ii. For R_n^2 , we assume that θ is increasing. For θ decreasing, similar computations as the one below can be made and the result holds true for θ as it has bounded variation on $[0, t_{n-1}]$. We thus observe that, for $k \geq 1$,

$$\theta\left(\frac{k-1}{n}\right) \leq n \int_{\frac{k-1}{n}}^{\frac{k}{n}} \theta(x) dx \leq \theta\left(\frac{k}{n}\right) .$$

Summing the previous inequalities and rearranging terms, we obtain

$$\left| \int_0^1 \theta(x) dx - \frac{1}{n} \sum_{k=0}^{n-1} \theta\left(\frac{k}{n}\right) \right| \leq \int_{1-\frac{1}{n}}^1 |\theta(x)| dx + \frac{|\theta(0)|}{n} + \frac{1}{n} \left| \theta\left(1 - \frac{1}{n}\right) \right| .$$

We then compute

$$\frac{1}{n} \left| \theta \left(1 - \frac{1}{n} \right) \right| \leq C \frac{1}{n^{\gamma(\beta^* - \beta + 1) + 1}} \leq \frac{K}{n},$$

and

$$\int_{1-\frac{1}{n}}^1 |\theta(x)| dx \leq C \int_{1-\frac{1}{n}}^1 (1-x)^{\gamma(\beta^* - \beta + 1) - 1} dx \leq \frac{K}{n},$$

which concludes the proof. \square

Using a similar development as in Lemma A.4, we deduce

Corollary A.5. *Under the same Assumptions of Lemma A.4, we have that there exists $K > 0$ such that*

$$(63) \quad \sum_{k=0}^{n-1} \psi(\hat{t}_k) \hat{h}_k^\ell = n^{-(\ell-1)} \left(T^{1/\gamma} \gamma \right)^{\ell-1} \int_0^T \psi(t) (T-t)^{\beta^*} dt + K n^{-\ell},$$

where \hat{t} and \hat{h} are defined in (18).

APPENDIX B. PROOF OF RESULTS ON SPARSE GRIDS

Proof of Proposition 4.1. The proof follows closely the arguments of Lemma 3.6 and Theorem 3.8 in [7]: The main difference is that we consider sparse grids with boundary points (corresponding to level zero in their description).

1. Note simply that the set

$$\mathcal{V} = \bigoplus_{p=0}^{\infty} \mathcal{V}_p(A)$$

coincides with $H^1(A)$ up to completeness with respect to the Sobolev norm $H^1(A)$ (cf. [7] equations (3.15)-(3.19), taking into account the boundary values). The sparse grid approximation is nothing else than a direct sum truncation. The error is dominated by the first level not included, and the control is deduced also for the non-zero boundary condition as in Theorem 3.8. in [7].

2. As shown in Lemma 3.6. in [7], the number of total nodes belonging to multi-levels with strictly positive entries, and having l_1 -norm at most p (denoted by $N_{p,d}^{>0, \text{Tot}}$) is found to be

$$N_{p,d}^{>0, \text{Total}} = \sum_{k=0}^{p-d} \binom{k+d-1}{d-1} 2^k.$$

We can then show, explicitly exposing the leading term, that

$$\begin{aligned} N_{p,d}^{>0, \text{Total}} &= \sum_{i=0}^{d-1} \binom{p}{i} (-1)^{d-1-i} 2^{p-i} + (-1)^d \\ &= 2^{p-d+1} \left(\frac{(p-d+1)^{d-1}}{(d-1)!} + O((p-d+1)^{d-2}) \right), \quad \text{for } p \geq d. \end{aligned}$$

Let us now consider all multi-levels. Let $N_{p,d}^{\text{Total}} = \# \{ \tilde{\mathbf{x}}_{\mathbf{l}, \mathbf{j}} : (\mathbf{l}, \mathbf{j}) \in \mathcal{I}_p(A) \}$ be the number of total nodes belonging to multi-levels having l_1 -norm at most p . We get

$$N_{p,d}^{\text{Total}} = \underbrace{2^d}_{\text{all zero level}} + \sum_{n_0=(d-p)^+}^{d-1} \binom{d}{n_0} 2^{n_0} N_{p, d-n_0}^{>0, \text{Total}}.$$

i.e., we find this quantity by choosing some zero-dimensions and then imposing on the other levels to have strictly positive entries, and having l_1 -norm at most p .

By similar arguments as in [7], we can explicitly exhibit the leading order term. Indeed, for $p > 2d$,

$$\begin{aligned}
N_{p,d}^{\text{Total}} &= 2^d + \sum_{n_0=0}^{d-1} \binom{d}{n_0} 2^{n_0} N_{p,d-n_0}^{>0,\text{Total}} \\
&= 2^d + \sum_{n_0=0}^{d-1} \sum_{i=0}^{d-n_0-1} \left(\binom{p}{i} (-1)^{d-n_0-1-i} 2^{p-i} + (-1)^{d-n_0} \right) \binom{d}{n_0} 2^{n_0} \\
&= 2^d + \sum_{n_0=0}^{d-1} \sum_{i=0}^{d-n_0-1} \left(\binom{p}{i} \binom{d}{n_0} (-1)^{d-1-i+n_0} 2^{p-i+n_0} \right) + \sum_{n_0=0}^{d-1} \sum_{i=0}^{d-n_0-1} (-1)^{d-n_0} \binom{d}{n_0} 2^{n_0} \\
&= 2^d + \sum_{i=0}^{d-1} \binom{p}{i} \sum_{n_0=0}^{d-i-1} \left(\binom{d}{n_0} (-1)^{d-1-i+n_0} 2^{p-i+n_0} \right) + (-1)^d \sum_{n_0=0}^{d-1} (d-n_0) \binom{d}{n_0} (-2)^{n_0} \\
&= \sum_{i=0}^{d-1} \binom{p}{i} \sum_{n_0=0}^{d-i-1} \left(\binom{d}{n_0} (-1)^{d-1-i+n_0} 2^{p-i+n_0} \right) + 2^d + d.
\end{aligned}$$

From which we deduce, that also in this case,

$$N_{p,d}^{>0,\text{Total}} = 2^{p-d+1} \left(\frac{(p-d+1)^{d-1}}{(d-1)!} + O((p-d+1)^{d-2}) \right), \quad \text{for } p > 2d.$$

□

REFERENCES

- [1] Robert Azencott. Densité des diffusions en temps petit: développements asymptotiques. In *Séminaire de Probabilités XVIII 1982/83*, pages 402–498. Springer Berlin Heidelberg, 1984.
- [2] Vlad Bally and Gilles Pagès. A quantization algorithm for solving multidimensional discrete-time optimal stopping problems. *Bernoulli*, 9(6):1003–1049, December 2003. Mathematical Reviews number (MathSciNet): MR2046816; Zentralblatt MATH identifier: 02072449.
- [3] Christian Bender and Jianfeng Zhang. Time discretization and Markovian iteration for coupled FBSDEs. *The Annals of Applied Probability*, 18(1):143–177, February 2008. Mathematical Reviews number (MathSciNet): MR2380895; Zentralblatt MATH identifier: 1142.65005.
- [4] Bruno Bouchard and Nizar Touzi. Discrete-time approximation and Monte-Carlo simulation of backward stochastic differential equations. *Stochastic Processes and their Applications*, 111(2):175–206, June 2004.
- [5] Philippe Briand, Bernard Delyon, and Jean Mémin. Donsker-type theorem for BSDEs. *Electronic Communications in Probability*, bf6:1–14 (electronic), 2001.
- [6] Philippe Briand and Céline Labart. Simulation of BSDEs by Wiener chaos expansion. *The Annals of Applied Probability*, 24(3):1129–1171, June 2014.
- [7] Hans-Joachim Bungartz and Michael Griebel. Sparse grids. *Acta Numerica*, 13(May):147, 2004.
- [8] J. Chassagneux and A. Richou. Numerical Stability Analysis of the Euler Scheme for BSDEs. *SIAM Journal on Numerical Analysis*, 53(2):1172–1193, January 2015.
- [9] Jean-François Chassagneux. Linear multistep schemes for BSDEs. *SIAM Journal on Numerical Analysis*, 52(6):2815–2836, 2014.
- [10] Jean-François Chassagneux and Adrien Richou. Rate of convergence for the discrete-time approximation of reflected BSDEs arising in switching problems. *arXiv preprint arXiv:1602.00015*, 2016.
- [11] Jean-François Chassagneux and Dan Crisan. Runge–Kutta schemes for backward stochastic differential equations. *The Annals of Applied Probability*, 24(2):679–720, April 2014.
- [12] Jean-François Chassagneux, François Delarue, and Dan Crisan. Numerical Method for McKean-Vlasov FBSDEs. *preprint*, 2017.
- [13] Paul-Eric Chaudru de Raynal and Camilo A. Garcia Trillos. A cubature based algorithm to solve decoupled McKean-Vlasov forward-backward stochastic differential equations. *Stochastic Processes and their Applications*, 125(6):2206–2255, 2015.

- [14] Dan Crisan and François Delarue. Sharp derivative bounds for solutions of degenerate semi-linear partial differential equations. *Journal of Functional Analysis*, 263(10):3024–3101, November 2012.
- [15] Dan Crisan and Terry Lyons. Minimal entropy approximations and optimal algorithms. *Monte Carlo Methods and Applications*, 8(4):343–355, 2002.
- [16] Dan Crisan and Konstantinos Manolarakis. Solving Backward Stochastic Differential Equations Using the Cubature Method: Application to Nonlinear Pricing. *SIAM Journal on Financial Mathematics*, 3(1):534–571, January 2012.
- [17] Dan Crisan and Konstantinos Manolarakis. Second order discretization of backward SDEs and simulation with the cubature method. *The Annals of Applied Probability*, 24(2):652–678, April 2014.
- [18] Dan Crisan, Konstantinos Manolarakis, and Colm Nee. Cubature methods and applications. In *Paris-Princeton Lectures on Mathematical Finance 2013*, pages 203–316. Springer, 2013.
- [19] Dan Crisan, Konstantinos Manolarakis, and Nizar Touzi. On the Monte Carlo simulation of BSDEs: An improvement on the Malliavin weights. *Stochastic Processes and their Applications*, 120(7):1133–1158, 2010.
- [20] François Delarue and Stéphane Menozzi. A forward-backward stochastic algorithm for quasi-linear PDEs. *The Annals of Applied Probability*, bf16(1):140–184, 2006.
- [21] Avner Friedman. *Partial Differential Equations of Parabolic Type*. Dover Publications, April 2013.
- [22] Emmanuel Gobet and Céline Labart. Error expansion for the discretization of backward stochastic differential equations. *Stochastic processes and their applications*, 117(7):803–829, 2007.
- [23] Emmanuel Gobet, Jean-Philippe Lemor, and Xavier Warin. A regression-based Monte Carlo method to solve backward stochastic differential equations. *The Annals of Applied Probability*, 15(3):2172–2202, August 2005. Mathematical Reviews number (MathSciNet): MR2152657; Zentralblatt MATH identifier: 1083.60047.
- [24] Emmanuel Gobet and Plamen Turkedjiev. Linear regression MDP scheme for discrete backward stochastic differential equations under general conditions. *Mathematics of Computation*, 85(299):1359–1391, 2016.
- [25] Gergely Gyurkó and Terry Lyons. Efficient and Practical Implementations of Cubature on Wiener Space. In Dan Crisan, editor, *Stochastic Analysis 2010*, pages 73–111. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.
- [26] Pierre Henry-Labordere, Xiaolu Tan, and Nizar Touzi. A numerical algorithm for a class of BSDEs via the branching process. *Stochastic Processes and their Applications*, 124(2):1112–1140, 2014.
- [27] Peter Kloeden and Eckhard Platen. *Numerical Solution of Stochastic Differential Equations*, volume 23 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1992.
- [28] Christian Litterer and Terry Lyons. High order recombination and an application to cubature on Wiener space. *The Annals of Applied Probability*, 22(4):1301–1327, 2012.
- [29] Terry Lyons and Nicolas Victoir. Cubature on Wiener space. *Proceedings of The Royal Society of London. Series A. Mathematical, Physical and Engineering Sciences*, 460(2041):169–198, 2004.
- [30] Colm Nee. *Sharp Gradient Bounds for the Diffusion Semigroup*. Imperial College London (University of London), 2011.
- [31] Gilles Pagès and Abass Sagna. Improved error bounds for quantization based numerical schemes for BSDE and nonlinear filtering. *ArXiv e-prints*, October 2015.
- [32] Étienne Pardoux and Shige Peng. Backward Stochastic Differential Equations and Quasilinear Parabolic Partial Differential Equations. *Stochastic Partial Differential Equations and Their Applications*, 176:200–217, 1992.
- [33] Denis Talay and Luciano Tubaro. Expansion of the Global Error for Numerical Schemes Solving Stochastic Differential Equations. *Stochastic Analysis and Applications*, 8(4):483–509, 1990.
- [34] Guannan Zhang, Max Gunzburger, and Weidong Zhao. A Sparse-Grid Method for Multi-Dimensional Backward Stochastic Differential Equations. *Journal of Computational Mathematics*, 31(3):221–248, May 2013.
- [35] Jianfeng Zhang. A numerical scheme for BSDEs. *The Annals of Applied Probability*, 14(1):459–488, 2004.